A study of statistical growth models and growth database.

VEKKATESWARLU, P.

A STUDY OF STATISTICAL GROWTH MODELS AND GROWTH DATABASE

P VENKATESWARLU

A thesis submitted to Sheffield Hallam University

in fulfillment of the requirements for the degree of

Master of Philosophy

Sponsoring Establishment
International Crops Research Institute for the Semi-Arid Tropics, Patancheru
Andhra Pradesh 502 324, India

July 1993

CONTENTS

A STUDY OF STATISTICAL GROWTH MODELS AND GROWTH DATABASE

P Venkateswarlu

Plant Physiologists and crop modelers in general are keen to use non-linear models for their work. However, usual practice remains restricted to the use of linear form of growth curves due to lack of proper methodology for the application of non-linear models.

This study is undertaken to review the statistical literature available on non-linear models, comparison of these models with data sets available.

A database management system known as GROWDAT for growth data has been developed for a systematic storage and retrieval of growth data collected from several experiments. A copy of the software will be given on request.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Dr G K Kanji, my first supervisor for his encouragement to undertake the work and for all his invaluable help during this study.

My sincere thanks and profound gratitude are due to Dr. Murari Singh, my second supervisor for his full support and guidance during my study.

I also thank several physiologists at the International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India, who have extended their valuable suggestions for the development of growth database.

Lastly, my thanks are due to my little daughters for patiently waiting for Daddy to come home for playing.

1. Introduction

A clear idea of development behaviour of plants is essential for human endeavour to improve them. Often plant scientists are interested in selecting an ideotype or a genotype with a desired growth behaviour, to cope up with various stress factors anticipated over a specific time interval and yet maintaining acceptable yield levels. In order to study the plant growth, experiments are conducted on desired genotypes of the plant and with a set of treatments applied on them, a sequence of response (such as height, diameter, drymatter etc.) are recorded over the growth period. Using the record of the response over time points, a functional form of the growth is evaluated. The parameters of the function are interpreted from biological or physiological points of view. These parameters are the characteristics of the plant genotype and the treatment (if any) applied on these genotypes. The selection of the genotype and the treatments is then done on the basis of the estimates of the parameters of the growth functions or the curves.

A number of models or functional forms for the plant growth will be reviewed in chapter 3 and the philosophy behind statistical modelling in chapter 2. It is quite essential to realize that in practice, we generate responses based on experiments conducted in field, green houses, plastic house, laboratories or incubators. These responses therefore have a component of experimental error. In order to make a valid and precise assessment of the responses at selected time points, it is essential

to follow principles of experimental designs advocated by Sir R.A.Fisher. Once the responses have been obtained they are better explained by models which incorporate stochastic components. Such models are called 'Statistical Models'. When the stochastic component is ignored, we have the deterministic models. The statistical models are relatively more realistic and if a suitable procedure for estimation of model parameters is followed, one can assess the reliability of the estimates.

A set of eight experiments from humid sub-humid trials of Forest/Fuelwood Research and Development Project, Bangkok, were conducted on *Acacia auriculiformis, A. mangium and Leucaena* trees. Summary of the growth behaviour of the genotypes and cutting management treatments on height, diameter at breast height and survival percentage are presented in Chapter 4.

The study of plant growth is an essential component of research in physiology. The data sets for such research are in the form of time series and differ from the data sets of those plant disciplines which require one time observations on experimental units (generally at final stage, e.g. yield of a crop at maturity). Due to this specialized nature of data set, an appropriate software for the data storage, editing, retrieval and analysis becomes necessary. In Chapter 5, Growth database developed in dBase III is presented. The instructions for using GROWDAT are presented in Appendix 2. This database is available from the author on request.

2 Statistical Modelling

2.1 Introduction

It is an approach and more than just a collection of techniques and models. As noted by Meadows (1980): " Statistical modelling should be considered a paradigm in understanding in a similar way to that in which systems dynamics or econometrics can be considered"

The following elements may be considered as essential to the statistical modelling approach

1. The inherent randomness of systems and in consequence all data will contain variability.
2. Importance is attached to data, and data should be handled consistently.
3. A wide range of modelling skills will be required in addition to statistical skills.
4. The main objective is gaining insight into the situation being studied.
5. Statistical techniques will be prominent in the above.

Statistical modelling is an approach that can be applied to any area. Although it is most readily applied to 'hard' scientific studies and it can also be applied to 'softer' problems.

2.2 Purpose and Types of Models

It is essential to understand the precise meaning of the following questions

Do I need a model?
What type of model do I need
What approach to modelling shall I take?
What are the steps in modelling?

in order to model any behaviour.

Models come in many different forms. These can be considered as a subset of symbolic models in Ackoff and Sasceni (1968) categories of Conceptual models, Iconic Models, Analogue models, and Symbolic models

A list of reasons for using models could be that they communicate fact and ideas; generate new ideas; predict behaviour; provide insight into the behaviour; clarify thinking etc.

Jeffers (1982) gives various reasons for using ecological models including an orderly and logical representation of the underlying relationships; a means of communication between different research works; and a synthesis of available information. Thus there are three main themes in the purposes of models

i. Communication
ii. Prediction
iii. Understanding

The emphasis placed on the three aspects will depend on the particular situation. A contrast between (i) and (ii) and (iii) is often set up.

Gilchrist (1984) describes a contrast between conceptual and empirical aspects. The conceptual approach uses "logical reasoning, 'known theory', to obtain a model". Whereas the empirical uses only the empirical evidence, the data. In practice a mixture of the two approaches is used, what Gilchrist calls the eclectric approach.

In relation to the two aspects ie., descriptive and explanatory models, is the idea of 'biologically meaningful parameters'. That is the parameters involved in a model should have some biological interpretation. The problem with these parameters is that the statistical estimates of such parameters often do not have 'nice' statistical properties, they are often complex non-linear functions of the natural mathematical parameterization of the model (Causton and Venus (1981)). Hunt (1982) argues caution on rejecting a model simply because its parameters cannot be given any general biological significance. Information is often not supplied by the parameters but through their derivatives. As Hunt states 'parameters are messengers of reality, not reality itself'. Gilchrist also considers the problem of reality, 'a model is only a limited, and possibly distorted, picture of reality'. " A model can be seen as truth insofar as it makes 'unhidden' aspects of the situation being modelled that were previously hidden". He goes on to say that the important aspect is the adequacy of the model to reveal some aspects of reality.

When one looks at a flexible function such as the Richards function the empirical-mechanistic divide seem even fuzzier. It is a mechanistic model because it comes from a certain differential equation model with possible mechanistic interpretations? Or is it just a useful empirical model? Obviously the fit of the model alone does not necessarily imply the underlying mechanism. There has to be an interpretation of the mechanism, but how satisfactory does it need to be? There are general ideas about growth behind the models considered but are these ideas good enough for any situation?

There is a second problem with flexible models, that of falsification. The concept of a hypothesis or model being falsifiable is central to the standard scientific approach. If a model is very flexible then it becomes difficult to falsify it, so it is always right. For some model such as splines this is not a problem because no interpretation is placed on them, they are purely smoothing functions. For some of the generalized logistic models there is an implicit mechanism behind them. How useful are these 'always right' models. This brings us to the next subject to examine, criteria for models.

2.3 Criteria for Models

What makes a good model? Randers (1980) listed the following desirable characteristics of a model

a. Insight generating capacity
b. Descriptive realism
c. Mode reproduction ability
d. Transparency
e. Relevance
f. Ease of enrichment
g. Fertility (or new ideas, experiments etc)
h. Formal correspondence with data
i. Point predictive ability

These criteria were primarily set up from a systems dynamics viewpoint. The above can be used to set up some general criteria which may be more in sympathy with statistical modelling.

A. Data Correspondence

Does the model make use of all the available data? Is its use of data consistent, making use of pre selected criteria to judge closeness of fit. Have regions of inadequate data fit been shown to have no substantial effect on the overall model?

B. Justifiability

Is the complexity of the model required? Can the model be falsified given the quantity of information available? If not, is the model sensitive to these conjectures?

C. Applicability

Can model predict behaviour if required? Is it relevant to the end user?

D. Insight

Does the model increase understanding of the modelled system? Does it indicate areas of inadequate understanding?

2.4 Methodologies for Modelling

The idea of statistical modelling as a subject has grown considerably in the last few years. Following are the steps to be considered for statistical modelling as suggested by Gilchrist (1984):

1. Identification

   Selecting the most suitable model. The identification may be based on ideas about the situation (conceptual), the data (empirical) or a combination of both.

2. Estimation and Fitting

   The parameters of the model are estimated using suitable criteria and the model fitted to the available data.

3. Validation

The validity of the model is considered, this can take place at various stages in model development.

4. Application

The use of the model, this will effect all stages of the modelling process.

5. Iteration

The above stages are not linear but the modeler will pass back and forth between them.

This approach works well in areas such as linear regression modelling and time series modelling in which you are dealing with a well defined family of models and selecting the best from that family. If a broader view is taken and the subjective nature of the modelling process is to be fully considered then a modified methodology is required.

Several problem solving methodologies have been developed for soft systems ie systems involving qualitative variables and subjective judgement. One such methodology is due to Checkland (1972). The essential stages are

(a) Analysis
(b) Root definition of relevant systems
(c) Conceptualization

(d) Comparison of definitions of possible changes
(e) Selections
(f) Design and implementation
(g) Appraisal

These stages can be seen as (i) obtaining information about current systems; (ii) reducing the system to its basic purposes; (iii) constructing models; (iv) comparing results of models with the situation to suggest changes which can then be selected and applied.

Using the ideas of this methodology a new methodology for statistical modelling can be developed.

(1) Conceptual Analysis

The object of this phase is to produce a conceptual or 'ideas' model of the situation to be examined. It will contain all possible relationships and their forms and the data available. The use of diagrams such as system maps and influence diagram are an important tool at this stage.

(2) Model Type Generation

Using the insight gained from (1) a number of possible types of model to be used are considered (eg stochastic differential equation models, linear regression models, non linear regression models).

(3) Model Building

At this stage the models for each type are constructed. Within this stage Gilchrist's methodology (1-3) can be used for each model type. A single model need not emerge from each type as there may be several competing models with little objective distinction between them.

(4) Comparison of Models

The models from (3) will be compared with respect to both direct application and generated insight. This comparison will primarily be a subjective comparison.

(5) Generation of Further Model Types

As a result of (4) improved models may be suggested.

(6) Application and Appraisal

The results of the above model will be used and critically evaluated.

The above methodology can be placed within the modelling-data collection circle and one of the applications may be data generation. It is hoped that such a methodology will lead to a more flexible approach to statistical modelling.

So far only the role of models has been elaborated to explain the real behaviour and

criteria guiding the models and conceptual steps in identifying the models. In the following chapter, emphasis is given on the models specific to the study of plant growth.

3. Plant Growth Models

3.1 Introduction

Linear models have dominated the statistical methods for investigating relationships not because such models are always the most appropriate but because the theory of fitting such models to data is very simple. The calculations involved in obtaining estimates of the parameters in linear models requires only the solution of a set of simple simultaneous equations.

In contrast more realistic forms of models which involve parameters in a non-linear fashion cannot be so simply fitted without the use of a computer. Some forms of nonlinear models were investigated before the development of modern computers, and the complicated methods of fitting them were devised later. However these models inevitably had little appeal to research workers and were not widely used in the past because of their complexity. With the availability of high speed computers the fitting of non-linear models should be no longer difficult than that of linear models. It is therefore important that the research biologist should be aware that there would be no difficulties in fitting these models.

3.2 Polynomial Growth Models

To understand why non-linear models should be useful, it is necessary to consider

why linear models are inadequate to biological situations. If we are considering a relationship between a stimulus variable, X, and a resulting yield variable, Y, then the three simplest forms of linear models are the straight line $Y = a + bx$, the quadratic form $Y = a + bx + cx^2$, and the cubic form $Y = a + bx + cx^2 + dx^3$. These are special cases of polynomial curves.

The straight line is obviously a very restricted relationship. Very few biological relationships are even approximately straight for a reasonable range of x values. The most common form of straight line relationship being, perhaps, the allometric relationship between the logarithm of weight of a plant or animal part and the logarithm of the whole plant or whole animal weight.

The quadratic model allows for curvature but is restricted in two critical ways. First, it is symmetric, the size of Y with increasing x to a maximum being of exactly the same form as the subsequent decline of Y with further increase in x. The symmetry can be avoided by considering not x but a power of x. $Y = a+b(x^{\circ})+c(x^{\circ})^2$ which is often a useful form of model, but is now a nonlinear parameter. The second disadvantage of a quadratic model is that the value of Y must become negative when x is either large or small and this will usually be biologically unreasonable. The cubic polynomial, and polynomials of yet higher degree overcome the disadvantage of symmetry but not those of producing unrealistically negative or very large values of Y for large or small values of x. In addition, they have a very rigid structure of

maximum and minimum values.

None of the curves in the polynomial family of models allows for a relationship which tends to an asymptotic level of Y as X becomes large, or relationships where Y is necessarily positive. In contrast, most of the nonlinear models in common use do allow such biologically realistic forms of behaviour. In addition, many of the commonly used nonlinear models can be derived from simple biological concepts which to the extent that they are appropriate, justify the use of the non-linear models.

3.3 Growth Data

Data on the growth of part, whole or groups of organism are collected in many areas of biological sciences. In agriculture or forestry the growth of crops/trees and animals is studied, in medicine it is the growth of individuals or the growth of tumors, and in botany the whole plant can be looked at or one can look at a particular part, for example a leaf or tiller.

Many biological investigations are also concerned with the growth or organisms with time. Extensive studies have been made of the growth of whole plants, or of the growth of the individual leaves, or of the growth of the animals. Qualitatively the growth of a biological organism can be thought of in four stages. Early growth, starting from a very small initial size, is relatively very rapid which can be expressed

quantitatively by saying that the rate of growth is proportional to the size of the organism. This form of growth is often called 'exponential growth'. If the rate of growth is exactly proportional to size then the size of the organism is described by the exponential function. The second stage is relatively less rapid, as more the energy of the organism is devoted to maintaining the current size. During this stage the growth of the organisms may well be approximated by the linear relationship. The third stage, the organism's growth diminishes further as a balance between the energy of the organism and the maintenance requirements is approached. The fourth and final stage of the growth is the antithesis of growth.

These above names were first used by Causton (1967); Rodford (1967) used the term 'dynamic' for what we call the 'functional' approach but terminology is relatively unimportant provided it is realized that one approach necessarily involves the use of fitted curves and the other does not. The classical approach in which the course of events is followed through a series of relatively infrequent, large harvests with much replication of measurements, the functional approach in which harvests, supplying data for curve fitting are smaller (less replication of measurements) but more frequent.

The data collection can be classified into two groups. The first is when a single entity is studied through time and measurements are taken at various times. The second is when a sample is taken from an experimental population at various times

in order to study the growth of the population as a whole. This is common in the study of crop growth when for example, in studying grass growth over a season a sample of grass given by a randomly placed quadrant would be harvested and weighted every week. One of the particular difficulties with the type of data, especially with the first type of collection, is that the observation will, in general, not be independent but correlated (Morgan, 1986). Glasbey (1979) considers five contributions to the errors of a model of the growth Ayrshire steel calves.

a. Variations in grid fill between weightings.

b. Seasonal variations and changes in diet

c. Illness

d. Errors in measuring procedure

e. Choice of wrong biometric form of the curve.

Of these b, c and e result in correlated errors.

In the case of sampling from a experimental population changes in the environment which would be common to the samples would also lead to the correlated error structure. Thus while standard regression modeling procedure involve the assumption of independent errors this will rarely be valid in the case of growth data.

Another problem is that, over a reasonable time span, the data is now linear. The

standard shape being that of a sigmoid curve, thus the most natural model form is a not linear model.

There are three basic approaches to this problem. One is to consider functions of the observations, usually some form of differences, and assure that these are, approximately independent. Some form of analysis is then carried out on these functions of data (Radford, 1967); Hunt, 1978,1982).

Alternatively, a second approach which will be called the statistical approach, relies on approximating the growth curve by a low order polynomial. The advantage of using a polynomial is that it is statistically simpler to deal with and a procedure involving a general error structure can be arrived at which only involves linear computations. However, it is rarely possible to assign a biological meaning to the parameters of the fitted model thus making interpretation difficult.

The third approach is to fit non-linear models whose parameters have a reasonable biological interpretation. Being non-linear these models involve greater problems in estimating the parameters than do the linear models. Numerical techniques have to be employed in order to minimize the resulting non-linear sums of squares function required by least squares procedure. Also little work has been done on fitting non-linear models with dependent errors (Morgan, 1986).

3.4 Fitting Growth Models

The following asymptotic functions were also examined with the data sets in addition to fitting exponential equations. In an asymptotic function, whatever its other properties, the form of its progression is governed by one characteristic feature: the value of the dependent variate more and more gradually ascends (or descends) to a plateau which it never quite meets. This plateau in the value of Y is known as the asymptote, or asymptotic value, and is the value predicted for Y when X is at infinity. Though this property of asymptotic functions is universal, it is not necessarily obvious in the form of every fitted progression. Unlike the polynomial functions, the asymptotic functions are statistically non-linear.

1. Monomolecular function
2. Logistic
3. Gompertz
4. Richards Function

3.4.1. Monomolecular function. The monomolecular function is described mathematically as

$$W(\text{or } \log_e W) = a(1-be^{-cT}) \qquad .. \ 1.1$$

where a is the asymptote, b a measure of the starting size of the system and c is a rate constant. The above equation is the basic function fitted to untransformed primary data. In asymptotic functions as in case of polynomials, asymptotic functions can also be subjected to logarithm transformation for the mathematical purpose of linearizing or simplifying the function to secure an easier method of fitting. Transforming both sides to logarithm produces

$$\log_e W = \log_e a + \log_e(1-be^{-cT}) \qquad .. 1.2$$

Here the parameter retain their original relationships vis-a-vis with the untransformed data and the growth function iteself becomes simpler with the changed properties (Causton, 1977). Transforming the left hand side only

$$\log_e w = a(1-be^{-cT}) \qquad .. 1.3$$

an equation which following the terminology adopted for polynomials by Causton (1970) may be termed the monomolecular exponential. This may be done in cases where the logarithms of primary data are homoscedastic and lie in a progression, the general shape of which corresponds broadly to that of the function. The structure and properties of the function are not altered by this transformation. It is merely applied to the logarithm of data instead of to their arithmetic values, back transformation of fitted data being necessary if the experimenter wishes for any

reason to perform subsequent comparisons on an untransformed basis. Values of $\log_e w$ may require the addition of constant to remove negatively and this must also be allowed for in a. There is an important difference between this procedure as that in which the whole function is transformed to logarithms as in equation 1.2. In this former case, if the experimenter is prepared to estimate and iterate around the asymptotic values, a, the basic monomolecular function written in the form of equation 1.2, a, may be linearized and fitted to data in the form

$$\log_e[1-(W/a)] = \log_e b - cT \qquad .. 1.4$$

and the same may be done for the monomolecular exponential equation 1.3

$$\log_e[1-(\log_e W/a)] = \log_e b - cT \qquad .. 1.5$$

Linearisation by logarithmic transformation is in equation 1.4 and 1.5, a mathematical device which brings the non-linear function within the bounds of simple regression methodology. But it also alters the statistical properties of the primary data.

The monomolecular function is one of the simplest of asymptotic functions. It has no point of inflection and its scope has a progression which is convex to the time axis, being proportional T the amount of growth yet to be made, a^{-w}. In this respect it can be useful in the same situations as one of the variants of the third order

polynomials though it differs from that function in that being genuinely asymptotic, it is unable to proceed into negative slope.

3.4.2. Logistic function. This function is, like the monomolecular, a three parameter function, is also known as the autocatalytic function and takes the form

$$W(\text{or } \log_e W) = a/(1+be^{-cT}) \qquad .. 1.6$$

where the bracketed term on the lefthand side refers to the dependent variate of the logistic exponential. The value of W (or $\log_e W$) at T= 0 is a/(1+b) and the function has a symmetrically placid point of inflection at T = $(\log_e b)/c$ and W for $\log_e W$ = a/2. After linearisation the functions may be fitted in the forms

$$\log_e[(a/W)-1] = \log_e b - cT \qquad .. 1.7$$

or

$$\log_e[(a/\log_e W)-1] = \log_e b - cT \qquad .. 1.8$$

For the logistic the slope are

$$\frac{dW}{dT} = abce^{-cT}/(1+be^{-cT})^2 \qquad .. 1.9$$

and

$$\frac{1}{W} \cdot \frac{dW}{dT} = bce^{-cT}/(1+be^{-cT}) \qquad .. \ 1.10$$

and for the logistic exponential

$$\frac{dW}{dT} = \frac{abce[a/(1+be^{-cT})-cT]}{(1+be^{-cT})^2} \qquad .. \ 1.11$$

and

$$\frac{1}{W} \cdot \frac{dW}{dT} = \frac{abce^{-cT}}{(1+be^{-cT})2} \qquad .. \ 1.12$$

The logistic equation has been used very extensively in the field of animal ecology for the modeling of change in numbers of individuals within a population (Solomon, 1976).

3.4.3. Gompertz Function. This function devised by Benjamin Gompertz in 1825, also has three parameters but these are arranged as a double exponent.

$$W(\text{or } \log_e W) = ae^{-be^{-ct}} \qquad .. \ 1.13$$

The value of W(or $\log_e$W) at T = 0 is ae^{-b}. Like the logistic the Gompertz's point of inflection occurs at T = $(\log_e b)/C$ but in the W (or $\log_e$W) dimension the curve

is asymmetrical with an inflection at a/e. The linear arrangements for the purposes of fitting are, for the Gompertz.

$$\log_e[\log_e(a/W)] = \log_e b - cT \qquad .. 1.14$$

and for the Gompertz exponential

$$\log_e[\log_e(a/\log_e W)] = \log_e b - cT \qquad .. 1.15$$

In both cases, the double logarithmic transformation is necessary to eliminate the double exponentiation present in equation in 1.13. The derivatives of the Gompertz are

$$\frac{dW}{dT} = abce^{-cT-be} \qquad .. 1.16$$

and

$$\frac{1}{W} \cdot \frac{dW}{dT} = bce^{-cT} \qquad .. 1.17$$

and for the Gompertz exponential

$$\frac{dW}{dT} = abce \qquad .. 1.18$$

and

$$\frac{1}{W} \cdot \frac{dW}{dT} = abce \qquad .. 1.19$$

Lengthy comparative discussions of the properties of the Gompertz curve, especially in relation to those of logistic, have been given by Winson (1932) and Richards (1969). The former warned that there seemed to be "no particular reason to expect that the Gompertz curve will show any wider range of fitting power than any other three constant S-shaped curve. The degree of skewness in the Gompertz curve is just as fixed as in the logistic and it is clear that to introduce a variable degree of skewness into a growth curve will require at least four constants. Of the Gompertz and logistic functions the latter wrote "since many (asymptotic) growth data are characterized by maximal rates somewhere within the range of a/3 to a/2, these can usually be fairly well accommodated by one or other of the two. Though this opinion did not prevent Richards from devising himself (in 1959) what has since come to be the most important of four parameter functions.

The majority of applications of the Gompertz function in plant growth analysis have been connected with the modeling of the growth of individual organs, particularly that of leaves.

3.4.4. Richard's Function. Unlike the three preceeding, this function, proposed by

F.J. Richards in 1959, has four parameters:

$$W(\text{or } \log_e W) = a(1 \pm e^{(b-cT)})^{-1/d} \qquad .. \ 1.20$$

and the second exponent, -1/d, involves the additional parameter, d. The term $e^{(b-cT)}$ is a more modern rearrangement of be^{-cT}, which was formerly used in this function in the same way as in the monomolecular, logistic and Gompertz.

Derivatives of the Richards function are

$$\frac{dW}{dT} = \frac{ace^{b-cT}}{d} . (1+e^{b-cT})^{-1(1/d+1)} \qquad .. \ 1.21$$

and

$$\frac{1}{W} . \frac{dW}{dT} = \frac{ce^{b-cT}}{d(1 \pm e^{b-cT})} \qquad .. \ 1.22$$

and for the Richards exponential

$$\frac{dW}{dT} = \frac{ace^{b-cT}}{d} . (1 \pm e^{b-cT})^{-(1/d+1)} . \qquad .. \ 1.23$$

$$e[a(1 \pm e^{b-cT})^{-1/d}]$$

and

$$\frac{1}{W}\cdot\frac{dW}{dT} = \frac{ace^{b-cT}}{d}\cdot(1 \pm e^{b-cT})^{-(1/d+1)} \quad .. \ 1.24$$

The previous three functions such as Monomolecular, logistic and Gompertz, have had linear forms which may be fitted if a value of a is first sought by trial and error. In the case of the Richards function, the originator's (1959) method of fitting was to select both a and b by this method. Subsequently, a long sequence of improvements on this approach has evolved (Nelder, 1961; Causton, 1969; Davies and Ku, 1977; Hadley, 1978) culminating into most modern medthod of fitting which enjoys automatically computed starting values, a separate treatment of a and d, stability of the method over wide variety of curve types. The Richards function represent a much wider class of growth phenomenon than over by some of the previous models which are its special cases.

Above curves can be easily fitted using a general statistical software, such as GENSTAT (A General Statistical Program) or SAS (Statistical Analytical Systems). For example, GENSTAT's FITCURVE commands has options for fitting ten standard non-linear curves. Some of these are the models discussed in this chapter.

This chapter presented various forms (linear, polynomial and non-linear including a major class of function known as 'Richards' function) for modeling plant growth. In reality, for a given plant genotype, many forms of the growth models discussed may not fit well. In the next chapter, screening of these models was carried out and

presented the functions fitting well to the growth behaviour of the trees selected under three cutting management methods.

4. Application of Growth Models

4.1 Introduction

Data from seven Humid and Sub-humid Zone Network Trials conducted by the Forestry/Fuelwood Research Development Project, Bangkok, have been used for the application of growth models. Experiments are coded as EXPT1 to EXPT7 for convenience.

Experimental Design

The experimental design is a randomized complete block in 3 replications, using a factorial arrangement of six genotypes or provenances x 3 management treatments. The two factors are labelled as Genotype and Cutting management treatments. Management treatment on each plot is applied to all live trees including trees in buffer rows. The total number of treatment combinations are 18. The genotypes considered in the experiment are nitrogen fixing forest trees. The characteristics of these forest trees are presented in Annexure 1 for the interest of the readers. They are *Acacia auriculiformis* PNG (papua new guinea), *Acacia auriculiformis* QLD (Queensland), *Leucaena hybrid* KX3, *Leucaena hybrid* K156, *Acacia mangium* PNG and *Acacia mangium* QLD. The three cutting management treatments are control, pollarding and pruning. The observations such as height, basal diameter, diameter

at breast height and survival percentage are recorded at an intervals of 6 months starting from 6 months upto 36 months. The procedure of recording various tree measurements are presented below.

Height

Measure the total height (Figure 1). If there are multiple stems, measure the tree from ground level to the top of its highest apical bud. If the site slopes significantly, measure the principal stem from the uphill of the tree. If the trees are bent over, straighten them if possible, so that the actual length of the stem is measured. Use a height stick or some type of marked, rigid pole to take height measurement.

Basal Diameter

Basal diameter is defined here as the diameter at 10cm above ground level (Figure 2). Take measurements on the principal stem (and other stems originating below 10cm). Mark with paint the point of measurement to ensure that the same point is measured subsequently. The measurement should be made with a metric diameter tape or vernier caliper. For multiple stemmed trees, the average diameter is calculated using the following formula:

$$da = SQRT(d1^2 + d2^2 + \ldots dn^2)$$

where da = average diameter and d1, d2 ...dn are the diameters of the stems. A

branching bole is considered a stem if its diameter is equal to or greater than 50% of the diameter of the principal bole at the same height.

Measurement should be made on the same trees used for height measurements at the ages of 6, 12, 18, 24, 30 and 36 months.

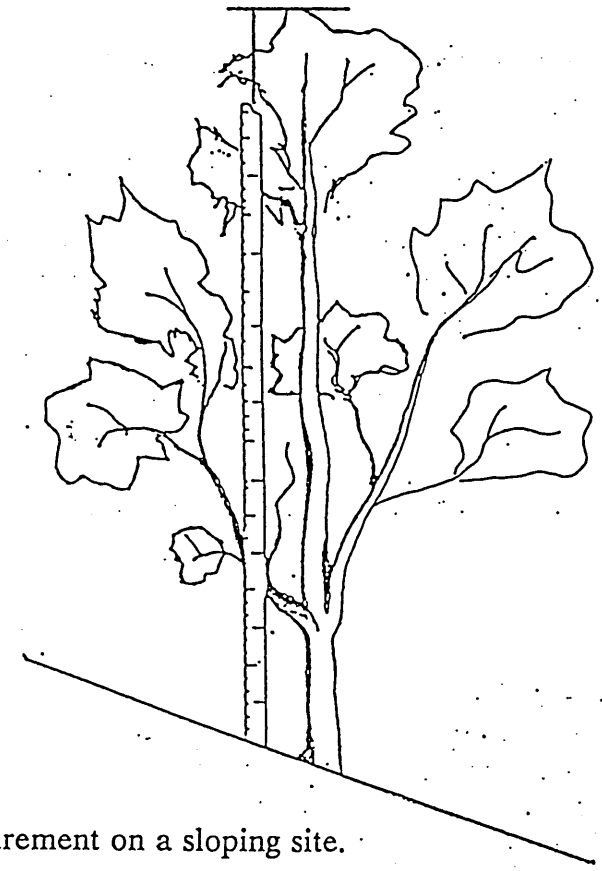
Diameter breast height (dbh)

Take dbh measurement at 1.3 meters above ground level (figure 2). For forking trees or trees with multiple leaders, take the dbh measurement on all stems if forking or multiple leaders originate less than 50cm above the ground. Mark with paint all the points of measurements. If there is an abnormality in the tree at 1.3 meters, measure the tree diameter at the point nearest this height which is representative of the stem's diameter.

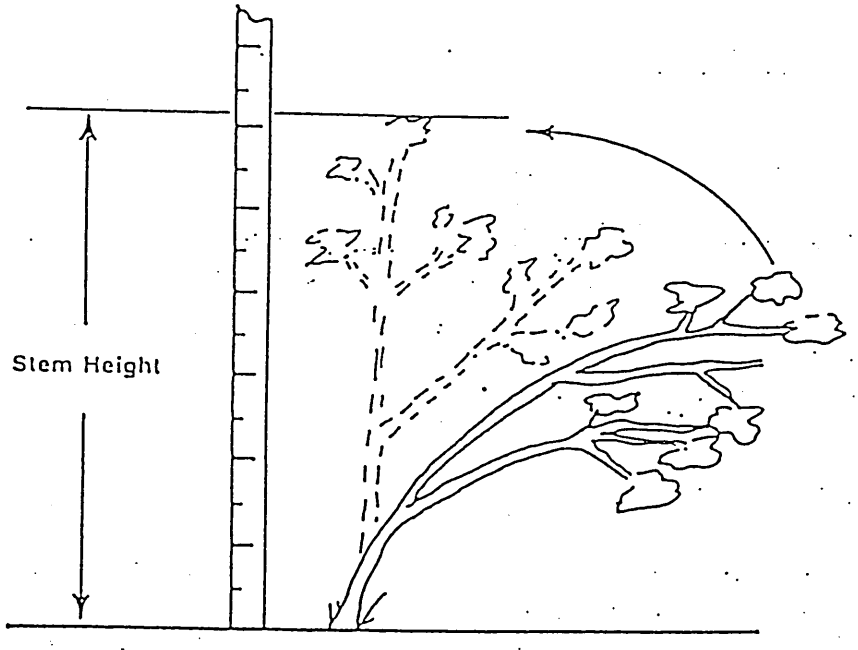
Survival

Survival is the number of live trees at the time of observation.

The common objective of the experiment was to assess the differences in the growth behaviour of various genotypes and pruning treatments for above four characters. In this chapter, we first study the variability due to genotype, cutting method and their interaction with time, explained as polynomial contrast. In many cases simple linear and quadratic contrasts do not explain the behaviour satisfactorily. Hence



(a) Height measurement on a sloping site.



(b) Height measurement of a 'bent' or leaning tree.

Figure 1. Measurement of tree height.

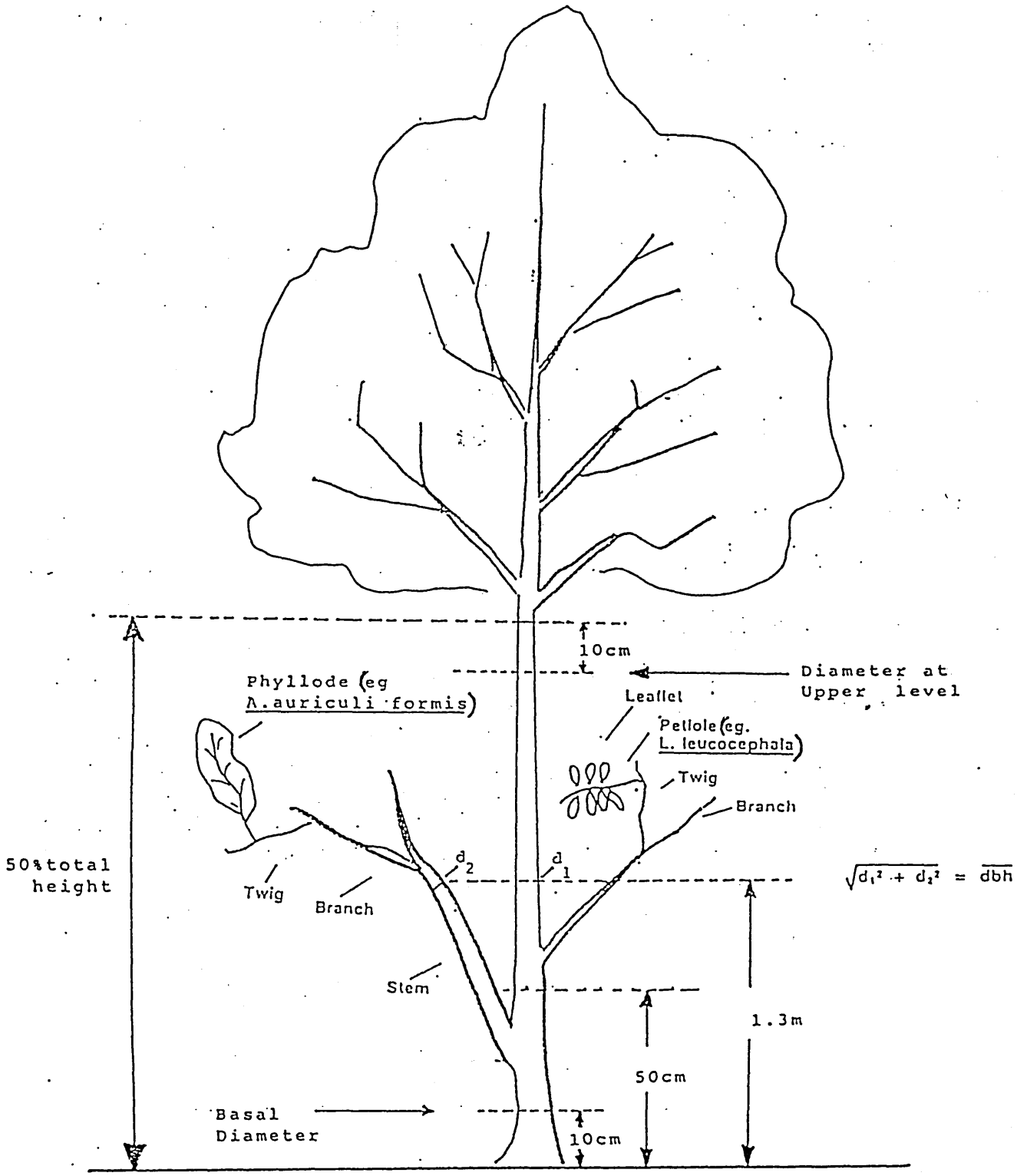


Figure 2. Diameter measurement guidelines.

attempt to fit growth curves was made and compared the related parameters of the growth models. For illustration, results from EXPT1 and EXPT2 are presented in the tables.

4.2 Analysis of Variance

The various induced sources were separated from the inherent sources of variation such as replication and experimental units (experimental error or residual) using the analysis of variance (skeleton anova shown in Table 1).

The ANOVA is of split block structure where genotype-cutting (main effects of genotypes, main effects of cuttings, interaction effects of genotypes and cutting) combinations are compared with an error different from that for interaction of genotype-cutting with time. The main effects of time (age) can not be tested for statistical significance since time can not be randomized in these experiments, but the interactions of genotype and cutting with time have valid test due to the (restricted) randomization of genotypes and cuttings. Summary of Analysis of variance for all the seven experiments are presented in table (Table 2a to 2g).

Significant interactions were noticed for genotype x cutting x age for all characters in EXPT1. Genotypic behaviour over time was also different. Cutting methods have significant influence on only height (Table 2a).

Table 1. Skeleton of Analysis of Variance

Source of Variation	df	SS	MSS
Rep	r-1		
Rep.Age Stratum			
Age	t-1		
Linear	1		
Quadratic	1		
Cubic	1		
Deviation	t-4		
Residual (a)	(r-1)(t-1)		
Rep.Genotype.Cutting Stratum			
Genotype	g-1		
Cutting	c-1		
Genotype.Cutting	(g-1)(c-1)		
Residual (b)	(r-1)(gc-1)		
Rep.Geno.Cutting.Age Stratum			
Genotype.Age	(g-1)(t-1)		
Genotype.Lin	g-1		
Genotype.Quad	g-1		
Genotype.Cubic	g-1		
Deviations	(g-1)(g-4)		
Cutting.Age	(c-1)(t-1)		
Cutting.Lin	c-1		
Cutting.Quad	c-1		
Cutting.Cubic	c-1		
Deviations	(c-1)(t-4)		
Genotype.Cutting.Age	(g-1)(c-1)(t-1)		
Genotype.Cutting.Lin	(g-1)(c-1)		
Genotype.Cutting.Quad	(g-1)(c-1)		
Genotype.Cutting.Cubic	(g-1)(c-1)		
Deviations	(g-1)(c-1)(t-4)		
Residual	(r-1)(gc-1)(t-1)		

where r = number of replications; t = number of time points;
g = number of genotypes; c = number of cutting managements.

Table 2a. Mean squares and probability levels in analysis of variance for four characters in EXPT1

Source of variation	d.f	Height	Prob.	DBH	Prob.	Basal	Prob.	Survival	Prob.
Rep	2	38.04		39.48		50.67		1377	
Rep.Age Stratum									
Age	5	134.32	<.001	134.96	<.001	224.19	<.001	4446	<.001
Rep.Genotype.Cutting Stratum									
Genotype	5	58.27	<.001	50.79	<.001	92.12	<.001	4994	<.001
Cutting	2	27.17	0.002	1.02	0.731	1.84	0.710	897	0.073
Genotype.Cutting	10	9.86	0.015	7.56	0.031	9.88	0.086	592	0.085
Residual	34	3.67		3.21		5.30		316	
Rep.Geno.Cutting.Age Stratum									
Genotype.Age	25	3.17	<.001	4.69	<.001	4.93	<.001	301	<.001
Genotype.Lin	5	13.16	<.001	18.89	<.001	19.82	<.001	1262	<.001
Genotype.Quad	5	1.58	0.004	0.03	0.998	0.84	0.069	50	0.266
Deviations	15	0.37	0.633	1.51	0.003	1.33	<.001	64	0.064
Cutting.Age	10	13.43	<.001	0.52	0.579	0.62	0.129	387	<.001
Cutting.Lin	2	37.22	<.001	0.62	0.366	1.03	0.079	1341	<.001
Cutting.Quad	2	8.43	<.001	0.28	0.633	1.35	0.037	315	<.001
Deviations	6	7.17	<.001	0.57	0.474	0.24	0.741	93	0.029
Genotype.Cutting.Age	50	1.40	<.001	0.91	0.033	0.81	<.001	111	<.001
Genotype.Cutting.Lin	10	4.77	<.001	1.93	<.001	2.67	<.001	270	<.001
Genotype.Cutting.Quad	10	0.81	0.058	0.32	0.875	0.48	0.302	192	<.001
Deviations	30	0.47	0.388	0.77	0.186	0.32	0.784	31	0.756
Residual	170	0.44		0.62		0.40		38	

Table 2b. Mean squares and probability levels in analysis of variance for four characters in EXPT2.

Source of variation	d.f	Height	Prob.	DBH	Prob.	Basal	Prob.	Survival	Prob.
Rep	2	18.69		11.52		8.21		99	
Rep.Age Stratum									
Age	5	270.13	<.001	63.31	<.001	223.27	<.001	3739	<.001
Rep.Genotype.Cutting Stratum									
Genotype	5	3.38	0.024	21.61	<.001	22.98	<.001	495	<.001
Cutting	2	45.94	<.001	5.30	0.036	6.02	0.034	306	0.005
Genotype.Cutting	10	1.44	0.287	1.33	0.526	1.57	0.486	255	<.001
Residual	34	1.13		1.44		1.61		48	
Rep.Geno.Cutting.Age Stratum									
Genotype.Age	25	1.56	<.001	1.10	<.001	1.52	<.001	108	<.001
Genotype.Lin	5	4.19	<.001	2.26	<.001	6.39	<.001	330	<.001
Genotype.Quad	5	2.10	<.001	0.91	<.001	0.17	0.548	65	0.035
Deviations	15	0.51	0.244	0.63	<.001	0.34	0.076	48	0.039
Cutting.Age	10	22.01	<.001	1.09	<.001	1.07	<.001	143	<.001
Cutting.Lin	2	65.40	<.001	1.66	<.001	3.53	<.001	487	<.001
Cutting.Quad	2	7.97	<.001	2.08	<.001	1.27	0.003	138	0.007
Deviations	6	12.22	<.001	0.31	0.019	0.18	0.541	29	0.347
Genotype.Cutting.Age	50	0.39	0.562	0.09	0.735	0.17	0.832	74	<.347
Genotype.Cutting.Lin	10	0.92	0.017	0.06	0.826	0.30	0.186	262	<.001
Genotype.Cutting.Quad	10	0.61	0.151	0.06	0.844	0.15	0.706	55	<.001
Deviations	30	0.14	0.999	0.11	0.349	0.13	0.942	18	0.028
Residual	170	0.41		0.10		0.21		27	

Table 2c. Mean squares and probability levels in analysis of variance for four characters in EXPT3.

Source of variation	d.f	Height	Prob.	DBH	Prob.	Basal	Prob.	Survival	Prob.
Rep	2	1.08		1.88		2.65		194	
Rep.Age Stratum									
Age	5	229.33	<.001	153.15	<.001	345.29	<.001	15860	<.001
Rep.Genotype.Cutting Stratum									
Genotype	5	12.51	<.001	17.09	0.022	29.05	0.004	2758	<.001
Cutting	2	5.37	0.084	0.20	0.965	0.50	0.929	434	0.366
Genotype.Cutting	10	2.54	0.287	4.53	0.618	4.33	0.773	107	0.987
Residual	34	2.01		5.57		6.81		420	
Rep.Geno.Cutting.Age Stratum									
Genotype.Age	25	1.76	<.001	2.04	<.001	2.15	<.001	301	<.001
Genotype.Lin	5	6.67	<.001	3.08	<.001	8.08	<.001	1262	<.001
Genotype.Quad	5	0.59	0.295	0.68	0.320	0.86	0.427	50	0.046
Deviations	15	0.52	0.376	2.20	<.001	0.60	0.744	64	<.001
Cutting.Age	10	3.80	<.001	0.52	0.511	0.68	0.652	387	0.213
Cutting.Lin	2	11.94	<.001	0.35	0.547	1.26	0.239	1341	0.234
Cutting.Quad	2	4.14	<.001	0.43	0.474	0.37	0.652	315	0.191
Deviations	6	0.97	0.064	0.65	0.341	0.58	0.675	93	0.315
Genotype.Cutting.Age	50	1.05	<.001	1.04	0.006	1.04	0.211	111	0.751
Genotype.Cutting.Lin	10	2.93	<.001	1.17	0.037	2.14	0.009	270	0.083
Genotype.Cutting.Quad	10	1.57	<.001	1.52	0.006	1.35	0.129	192	0.242
Deviations	30	0.25	0.982	0.74	0.190	0.56	0.921	31	0.991
Residual	170	0.48		0.57		0.87		38	

Table 2d. Mean squares and probability levels in analysis of variance for four characters in EXPT4.

Source of variation	d.f	Height	Prob.	DBH	Prob.	Basal	Prob.	Survival	Prob.
Rep	2	18.34		2.77		4.32		433	
Rep.Age Stratum									
Age	5	431.81	<.001	28.95	<.001	311.32	<.001	7147	<.001
Rep.Genotype.Cutting Stratum									
Genotype	5	11.58	0.002	28.95	<.001	16.03	0.006	1385	<.001
Cutting	2	86.59	<.001	20.51	0.008	26.04	0.004	902	0.073
Genotype.Cutting	10	1.22	0.867	2.84	0.658	2.98	0.679	238	0.085
Residual	34	2.38		3.70		4.01		374	
Rep.Geno.Cutting.Age Stratum									
Genotype.Age	25	2.70	<.001	0.98	<.001	1.37	<.001	205	<.001
Genotype.Lin	5	9.76	<.001	1.52	0.002	4.44	<.001	306	0.002
Genotype.Quad	5	0.89	0.234	0.79	0.064	1.06	0.078	254	0.006
Deviations	15	0.95	0.124	0.81	0.023	0.46	0.588	156	0.012
Cutting.Age	10	37.53	<.001	6.76	<.001	6.23	<.001	767	<.001
Cutting.Lin	2	118.31	<.001	12.31	<.001	25.06	<.001	2359	<.001
Cutting.Quad	2	34.42	<.001	13.27	<.001	3.78	0.001	366	0.008
Deviations	6	11.64	<.001	0.75	0.097	0.77	0.194	370	<.001
Genotype.Cutting.Age	50	0.78	0.191	0.44	0.234	0.77	0.036	119	0.014
Genotype.Cutting.Lin	10	1.92	0.002	0.49	0.213	2.19	<.001	174	0.013
Genotype.Cutting.Quad	10	1.26	0.043	0.26	0.701	0.30	0.825	149	0.035
Deviations	30	0.24	0.999	0.49	0.166	0.45	0.660	91	0.214
Residual	170	0.65		0.37		0.52		74	

Table 2e. Mean squares and probability levels in analysis of variance for four characters in EXPT5.

Source of variation	d.f	Height	Prob.	DBH	Prob.	Basal	Prob.	Survival	Prob.
Rep	2	2.94		1.28		0.71		1377	
Rep.Age Stratum									
Age	5	254.33	<.001	142.61	<.001	1585.90	<.001	4446	<.001
Rep.Genotype.Cutting Stratum									
Genotype	5	9.94	0.006	18.94	<.001	23.54	<.001	4994	<.001
Cutting	2	0.90	0.669	1.81	0.401	2.18	0.370	897	0.073
Genotype.Cutting	10	4.43	0.082	3.75	0.091	5.60	0.027	592	0.085
Residual	34	2.20		1.92		2.12		316	
Rep.Geno.Cutting.Age Stratum									
Genotype.Age	25	0.59	0.044	1.54	<.001	1.15	<.001	301	<.001
Genotype.Lin	5	0.99	0.016	3.74	<.001	3.03	<.001	1262	<.001
Genotype.Quad	5	0.64	0.089	0.30	0.652	0.21	0.670	50	0.266
Deviations	15	0.13	0.793	0.59	0.315	0.20	0.679	64	0.064
Cutting.Age	10	0.25	0.550	0.32	0.690	0.02	0.999	387	<.001
Cutting.Lin	2	0.66	0.122	0.32	0.516	0.04	0.878	1341	<.001
Cutting.Quad	2	0.09	0.726	0.29	0.554	0.00	0.994	315	<.001
Deviations	6	0.01	0.993	0.33	0.507	0.01	0.955	93	0.029
Genotype.Cutting.Age	50	0.09	0.999	0.45	0.572	0.22	0.883	111	<.001
Genotype.Cutting.Lin	10	0.21	0.695	1.15	0.024	0.55	0.149	270	<.001
Genotype.Cutting.Quad	10	0.03	0.999	0.14	0.964	0.05	0.995	192	<.001
Deviations	30	0.03	0.998	0.06	0.999	0.08	0.985	31	0.756
Residual	170	0.31		0.49		0.35		38	

Table 2f. Mean squares and probability levels in analysis of variance for four characters in EXPT6.

Source of variation	d.f	Height	Prob.	DBH	Prob.	Basal	Prob.	Survival	Prob.
Rep	2	17.44		17.62		14.98		75	
Rep.Age Stratum									
Age	5	119.62	<.001	106.34	<.001	101.18	<.001	4731	<.001
Rep.Genotype.Cutting Stratum									
Genotype	5	14.24	<.001	5.46	0.026	2.83	0.099	2669	<.001
Cutting	2	40.15	<.001	0.95	0.605	1.56	0.337	253	0.429
Genotype.Cutting	10	1.91	0.254	3.83	0.056	4.35	0.006	185	0.775
Residual	34	1.44		1.85		1.39		292	
Rep.Geno.Cutting.Age Stratum									
Genotype.Age	25	0.45	<.001	1.31	<.001	0.38	0.002	282	<.001
Genotype.Lin	5	1.24	<.001	3.43	<.001	0.72	0.001	1069	<.001
Genotype.Quad	5	0.38	0.026	2.05	0.001	0.65	0.003	49	0.311
Deviations	15	0.22	0.110	0.37	0.724	0.18	0.413	97	0.004
Cutting.Age	10	26.37	<.001	2.89	<.001	0.90	<.001	333	<.001
Cutting.Lin	2	81.19	<.001	6.99	<.001	0.63	0.029	1138	<.001
Cutting.Quad	2	16.75	<.001	4.20	<.001	1.06	0.003	367	<.001
Deviations	6	11.29	<.001	1.08	0.045	0.93	<.001	53	0.266
Genotype.Cutting.Age	50	0.11	0.814	0.61	0.167	0.26	0.031	111	<.001
Genotype.Cutting.Lin	10	0.19	0.196	1.21	0.009	0.72	<.001	291	<.001
Genotype.Cutting.Quad	10	0.15	0.383	0.66	0.212	0.17	0.444	146	<.001
Deviations	30	0.08	0.980	0.39	0.780	0.13	0.799	39	0.535
Residual	170	0.14		0.49		0.17		41	

Table 2g. Mean squares and probability levels in analysis of variance for four characters in EXPT7.

Source of variation	d.f	Height	Prob.	DBH	Prob.	Basal	Prob.	Survival	Prob.
Rep	2	2.39		1.57		2.27		1142	
Rep.Age Stratum									
Age	5	33.10	<.001	35.88	<.001	124.70	<.001	5981	<.001
Rep.Genotype.Cutting Stratum									
Genotype	5	3.41	0.002	6.04	0.009	12.55	<.001	4637	<.001
Cutting	2	0.78	0.241	2.18	0.190	0.48	0.717	2	0.994
Genotype.Cutting	10	0.50	0.467	3.23	0.043	1.37	0.468	720	0.010
Residual	34	0.51		1.22		1.41		190	
Rep.Geno.Cutting.Age Stratum									
Genotype.Age	25	0.44	<.001	1.00	0.010	1.55	<.001	11	<.001
Genotype.Lin	5	0.44	0.016	0.72	0.169	4.73	<.001	42	<.001
Genotype.Quad	5	0.53	0.006	0.27	0.592	1.14	0.033	3	0.065
Deviations	15	0.41	0.001	1.51	0.003	0.63	0.109	3	<.001
Cutting.Age	10	0.12	0.419	0.24	0.805	0.74	0.048	1	0.857
Cutting.Lin	2	0.06	0.610	0.36	0.423	1.00	0.075	0	0.942
Cutting.Quad	2	0.43	0.032	0.19	0.626	1.77	0.011	0	0.988
Deviations	6	0.05	0.890	0.19	0.766	0.29	0.584	1	0.513
Genotype.Cutting.Age	50	0.10	0.685	0.29	0.839	0.43	0.326	2	0.004
Genotype.Cutting.Lin	10	0.23	0.083	0.37	0.517	0.92	0.031	5	<.001
Genotype.Cutting.Quad	10	0.05	0.853	0.16	0.889	0.27	0.670	2	0.095
Deviations	30	0.08	0.865	0.32	0.678	0.32	0.654	1	0.540
Residual	170	0.12		0.42		0.38			

Results from Table 2b indicate that there is no significant interaction of genotype x cutting x age present for height, basal diameter and survival. However the genotype x age and cutting x age are significant.

The results from Table 2c to 2g indicate that the cutting management have significant influence over time on height than any other characters.

4.3 Fitting Growth Models

Examination of interactions Genotype x Cutting x Age, Genotype x Age, Cutting x Age and their partitioning into polynomial contrasts indicates whether different growth models be fitted for each genotype-cutting combination, each genotype or each cutting separately. If genotype x cutting x age interaction is significant, then we need to fit separate model for each genotype-cutting combination. If it is not significant, then we need to examine genotype x age interaction and cutting x age interactions. If genotype x age interaction is significant, we should fit separate curve for each genotype otherwise the genotypic growth curves would be parallel. Similarly, if cutting x age interaction is significant, then there is a need of fitting different curves to different cuttings. In absence of such an interaction, the growth curves for the cutting would be parallel.

Illustration of few cases of growth models fitted for genotypes, cuttings and their combinations are given below.

4.3.1 Exponential curves over genotype and cutting combinations in EXPT1

A close examination of genotype x cutting x age interaction in data set from EXPT1 indicates that growth behaviour of genotype-cutting combinations are different for all the four variables measured. We fitted then for each genotype-cutting combinations an exponential curve

$$Y = a + br^t$$

where 'r' is a shape parameter, b slope parameter and a+b is the intercept (at t=0), tested for differences in a', b' and r' over the genotype-cuttings combination. Genotype-cutting combinations are denoted by GenoCut (at 6x3=18 levels) and age by Age. GENSTAT5 commands for analysis on variable Y are

```
MODEL    Y
TERMS    Age*GenoCut
FITCURVE   Age        " FITCURVE has exponential by default"
ADD  GenoCut
ADD  Age.GenoCut
ADD [Print = m,s,e,a; nonl=s]
```

This resulted in the following accumulated analysis of variance.

Table 3. Accumulated analysis of variance for exponential curves fitting.

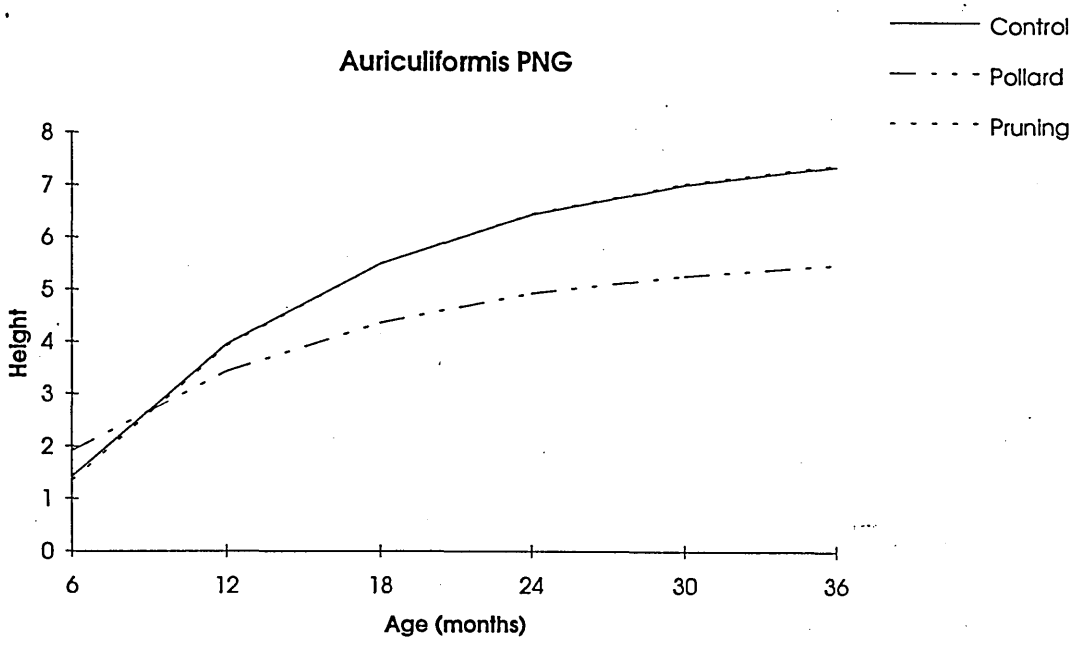
Source (Change)	df	Mean Squares			
		Height	DBH	Basal	Survival
+ Age	2	106.51	103.61	160.31	3382.6
+GenoCut	17	8.71	6.58	11.04	641.1
+Age.GenoCut	17	2.83	2.27^{ns}	2.68^{ns}	232.1
+separate nonlinear parameters	17	0.90^{ns}	0.10^{ns}	0.28^{ns}	52.6
Residual	54	0.78	0.61	1.10	27.10

ns - not significant at 5% level of significance.

Table 3 indicates that the non-linear parameter r is same over genotype-cutting combinations as the differences in "non-linear parameters" are not significant for all variables (at 5%) except survival (which was not significant at 1%). Estimates of these growth curve parameters for all characters are presented in Table 4. These models provide quite satisfactory fit to the data explaining more than 80% of the variability explained per experimental unit. Thus the exponential models with these coefficients can be used to predict the behaviour of the trees. The estimated lines from table 4 for genotypes are presented in figures 3-8 for height and diameter at breast height only.

Table 4. Exponential growth model parameter estimates from EXPT1 for different characters.

Genotype	Cutting	P	Estimates			
			Height	DBH	Basal	Survival
	Common	r	0.921	1.014	1.074	1.033
Auri-PNG	Control	b	-10.645	9.114	0.482	-0.598
		a	7.915	-8.301	3.030	98.710
	Pollard	b	-6.398	9.614	0.576	-24.736
		a	5.816	-9.155	2.453	135.897
	Pruning	b	-10.829	10.099	0.522	-0.968
		a	7.954	-9.557	2.812	101.38
Auri-QLD	Control	b	-11.110	11.018	0.676	-1.102
		a	7.929	-10.857	2.038	101.775
	Pollard	b	-4.676	9.163	0.506	-13.668
		a	5.282	-8.336	2.872	121.005
	Pruning	b	-13.848	13.938	0.744	-1.418
		a	9.190	-14.307	1.759	100.943
Leucaena Drive 156	Control	b	-6.535	4.955	0.362	-13.157
		a	4.984	-4.185	1.294	98.367
	Pollard	b	-3.779	3.118	0.266	-9.842
		a	3.856	-2.138	1.696	105.470
	Pruning	b	-2.768	1.969	0.119	-18.223
		a	3.456	-0.902	1.858	126.131
Hybr-KX3	Control	b	-3.018	2.632	0.189	-19.696
		a	2.997	-1.923	1.535	119.136
	Pollard	b	-2.942	6.156	0.267	-27.688
		a	3.083	-5.842	1.626	125.471
	Pruning	b	-5.394	4.581	0.360	-23.244
		a	4.752	-3.499	1.876	121.078
Mang-PNG	Control	b	-13.752	11.262	0.668	-2.954
		a	9.800	-9.503	3.480	103.132
	Pollard	b	-5.259	7.221	0.482	-15.837
		a	4.536	-5.993	2.111	123.478
	Pruning	b	-8.911	8.197	0.459	-6.317
		a	6.466	-7.529	2.297	108.632
Mang-QLD	Control	b	-7.376	5.670	0.396	-6.142
		a	5.084	-4.820	1.795	106.873
	Pollard	b	-4.014	5.994	0.314	-11.639
		a	4.397	-4.384	3.283	117.660
	Pruning	b	-8.998	7.353	0.487	-4.999
		a	5.827	-6.926	1.570	101.905
R^2 %			81.5	86.5	84.3	85.3



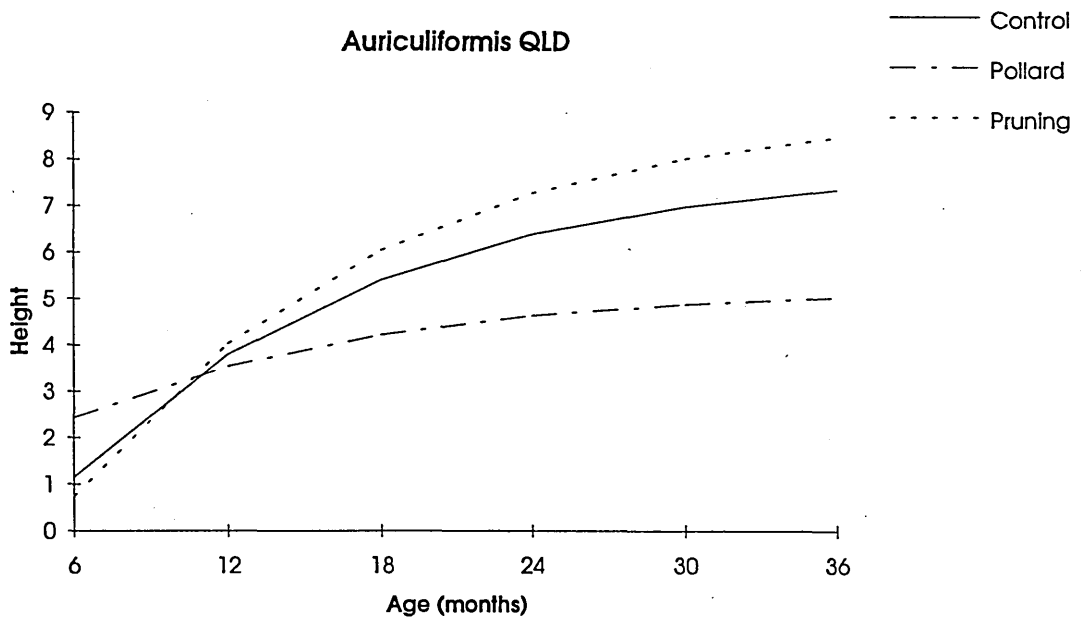
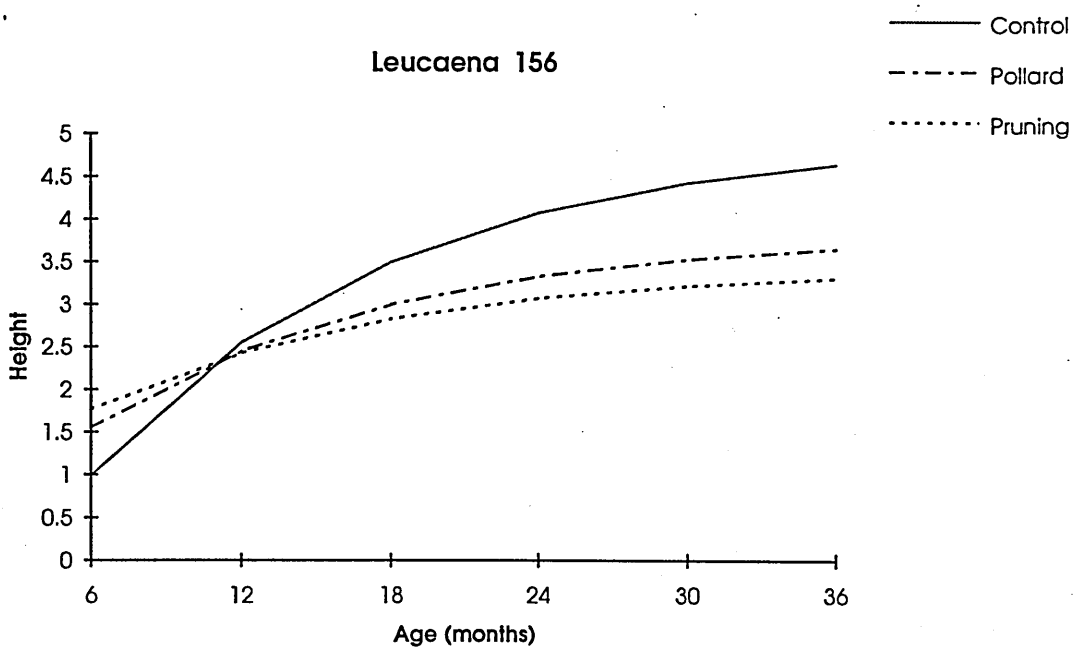


Figure 3. Height of Acacia auriculiformis at 6-36 months for three cutting management treatments.



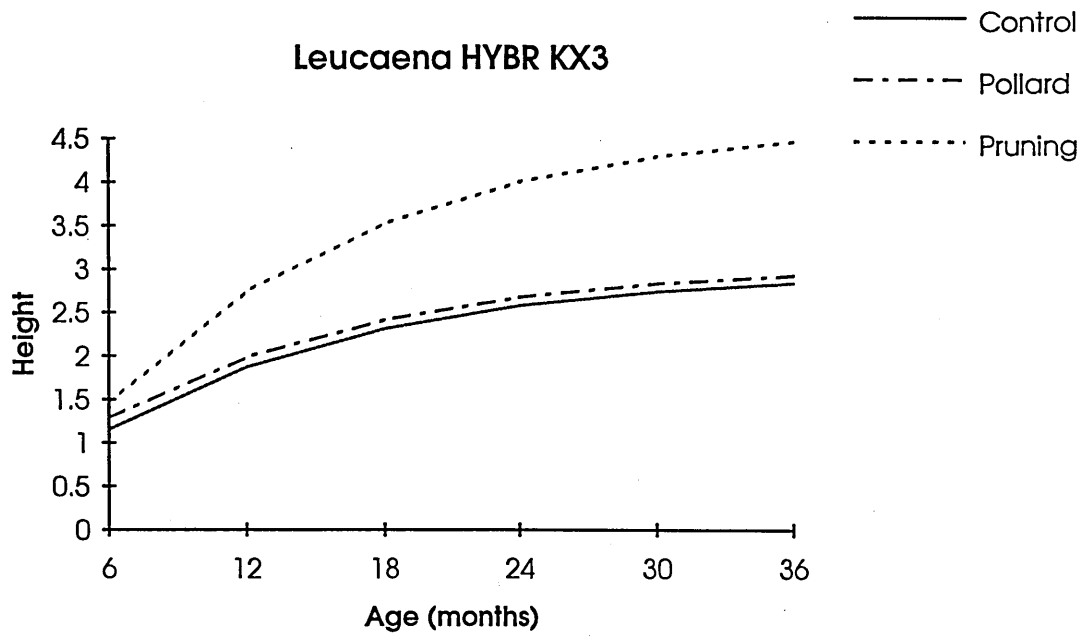
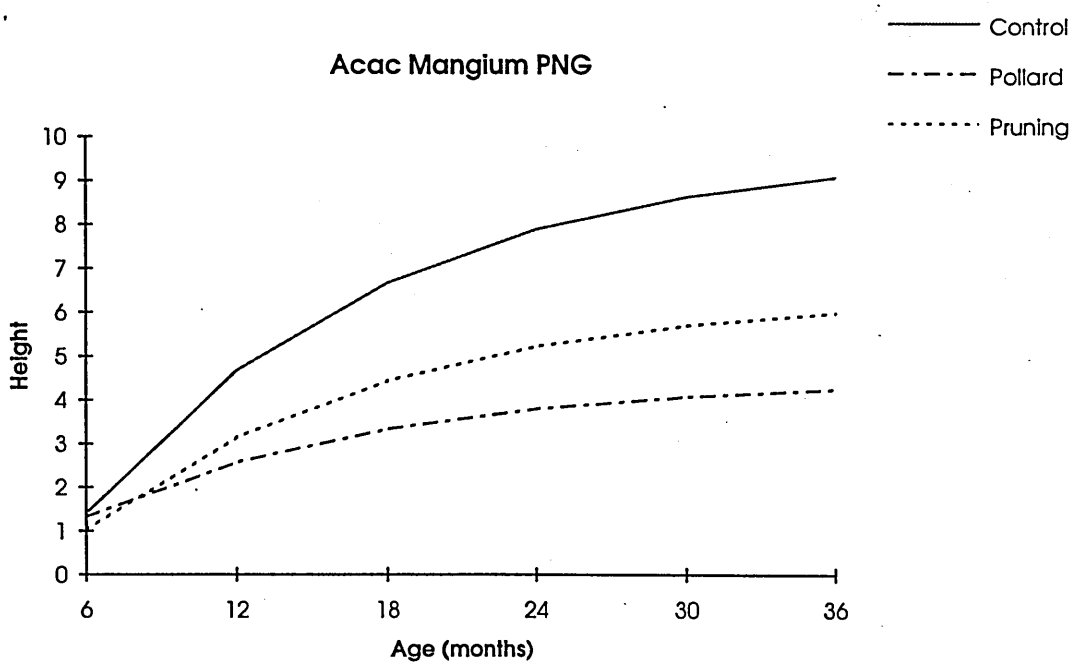


Figure 4. Height of Leucaena spp. at 6-36 months for three cutting management treatments



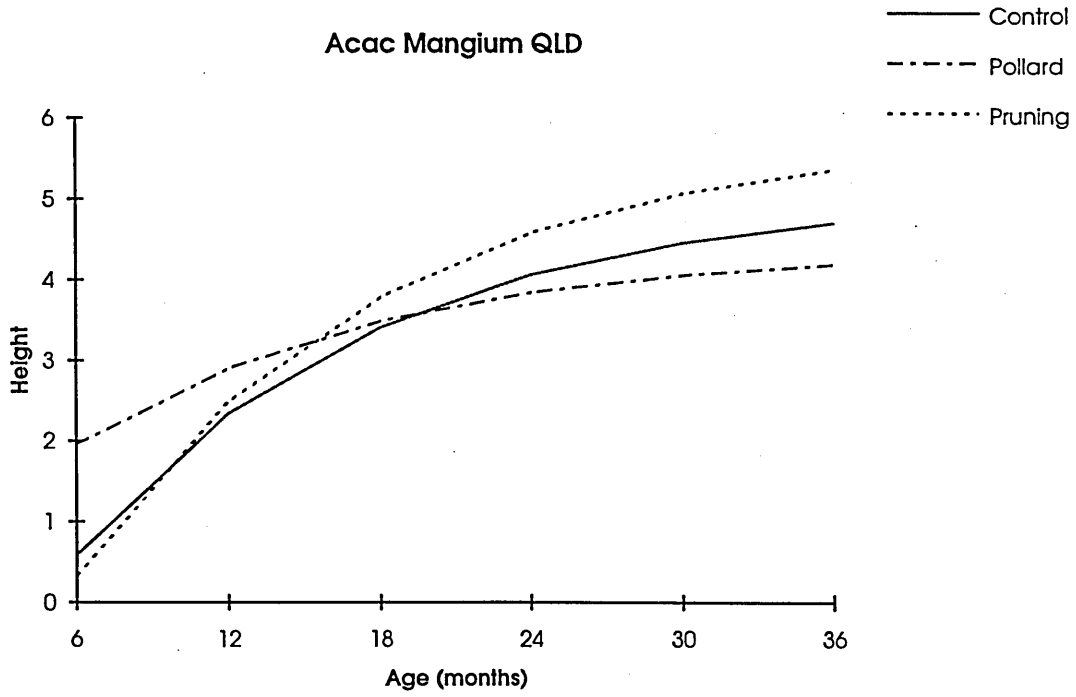
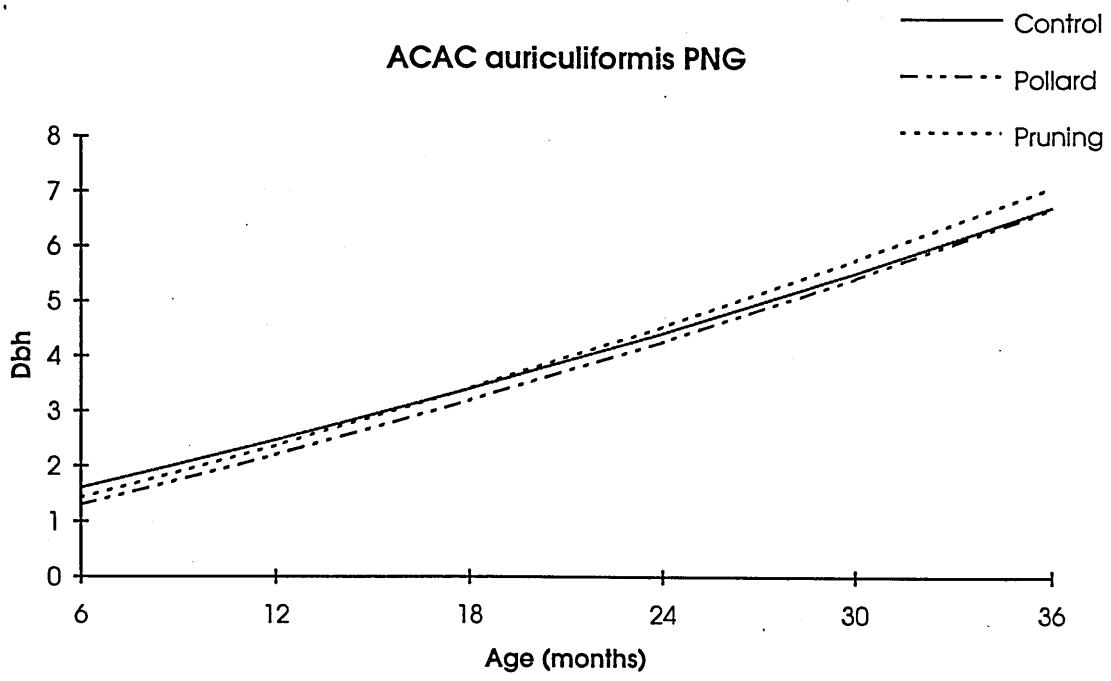


Figure 5. Height of Acacia mangium at 6-36 months for three cutting management treatments.



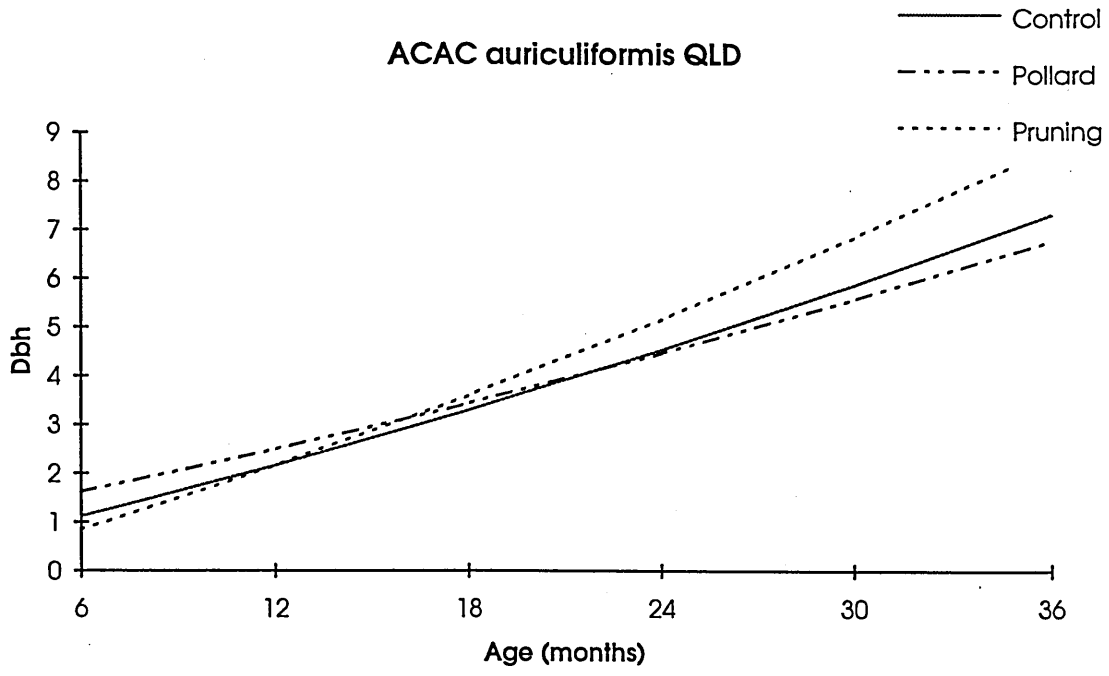
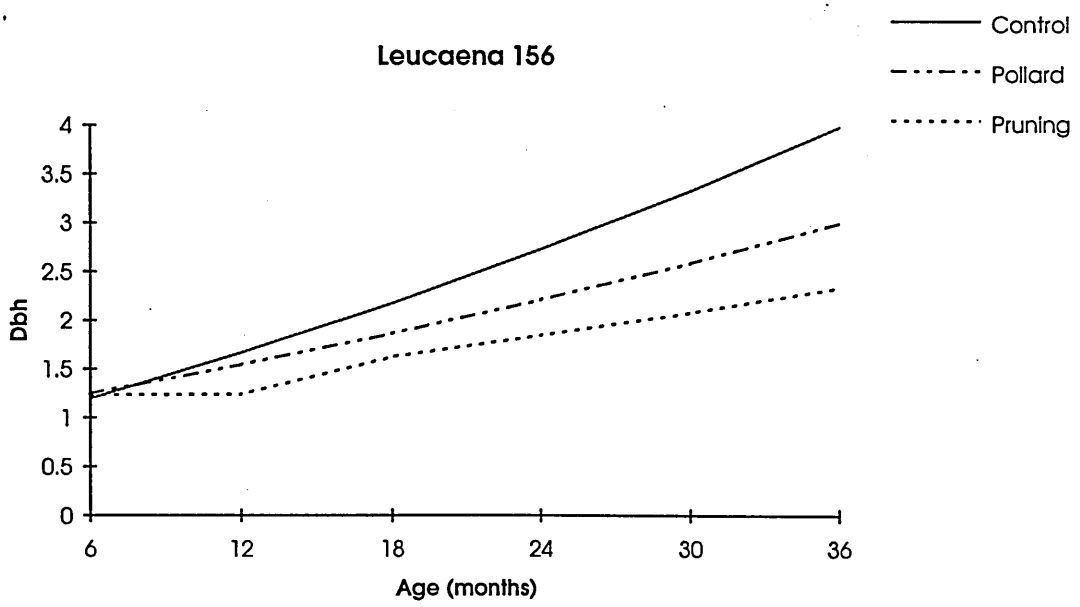


Figure 6. Diameter at breast height for three cutting treatments from equations at Table 4.



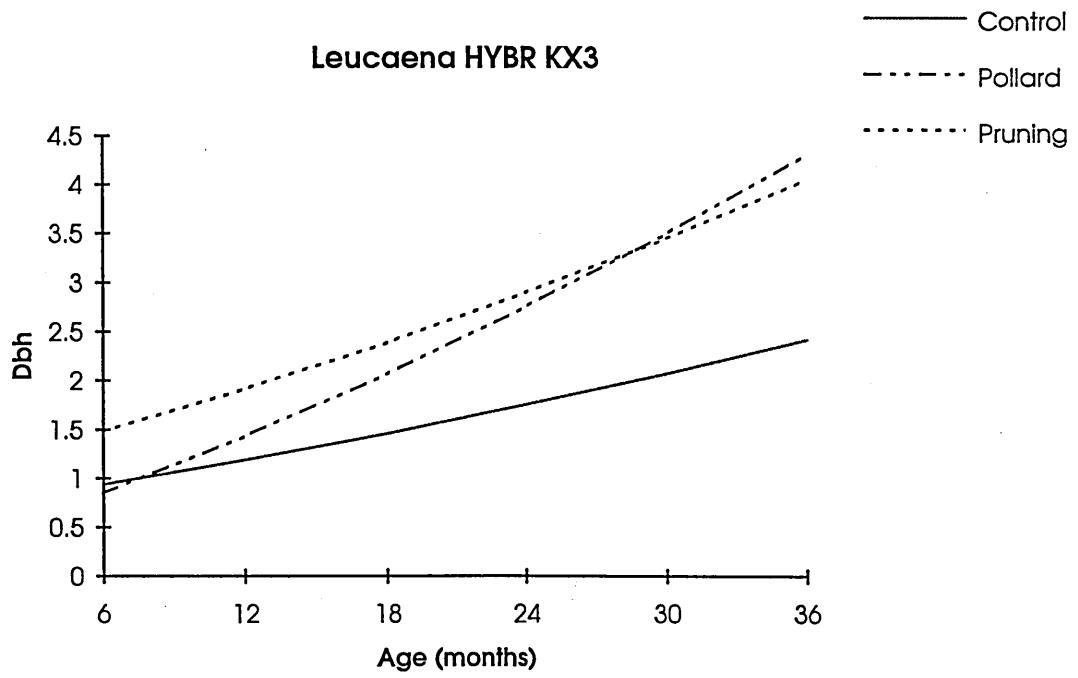
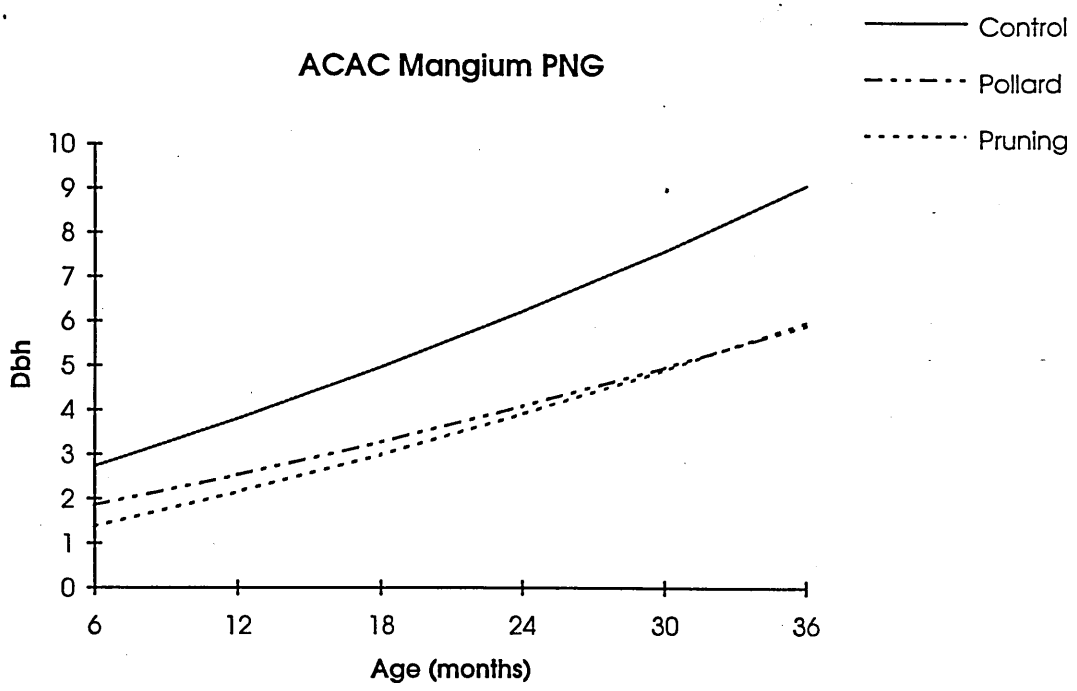


Figure 7. Diameter at breast height for three cutting management treatments for equation from Table 4.



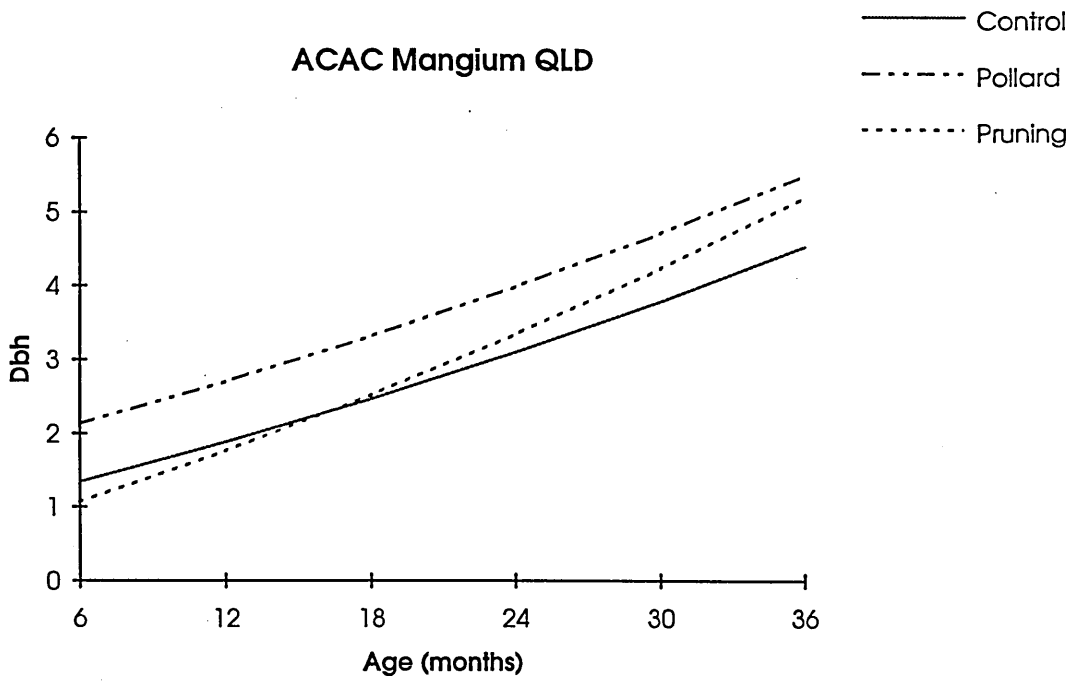


Figure 8. Diameter at breast height for three cutting management treatments for equations from Table 4.

4.3.2 Exponential growth models for Genotypes and Cuttings in EXPT2.

Let us now examine the genotype x cutting x age interaction for data from EXPT2. We note that this interaction is not significant for height and diameters. However, the interactions genotype x age and cutting x age are significant. In this case, it becomes worthwhile to fit the growth models for each genotype and for each cutting treatment. We present a rather selective models for genotypes and cuttings. Exponential growth model

$$Y = a + br^t$$

t is time in months have been fitted for height, diameters. The effect of treatment (genotype/cutting) dependent parameters on the variability in the response are presented in table 5.

Table 5. Accumulated analysis of variance when fitting exponential growth models on genotypes in EXPT2.

Source (change)	d.f	Mean Squares		
		Height	DBH	Basal
+ Age	2	71.23	13.87	61.44
+ Genotype	5	0.38ns	2.00	2.55
+ Age.Genotype	5	0.24ns	0.43	0.68
+ Separate non-linear	5	0.53ns	0.01ns	0.05ns
Residual	18	0.45	0.05	0.096

ns - not significant (at 5% level).

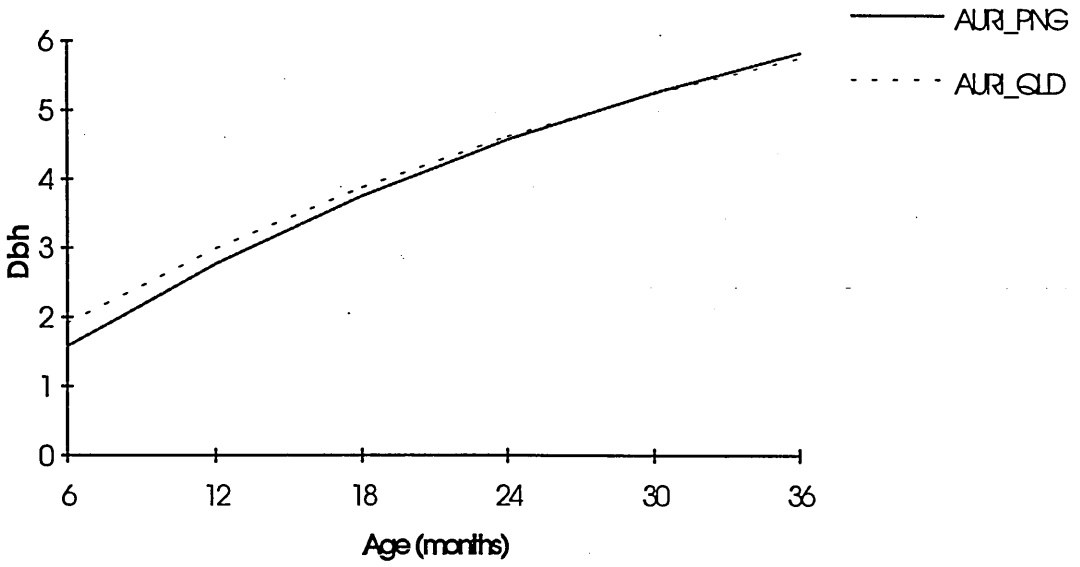
From the above table, it is clear that all the growth model parameters (a, b, r) are same over the six genotypes for height, while only r is same for diameters where the other two parameters a and b vary with genotypes. The estimates of the growth model parameters are presented in table 6 and the graphs are presented in figures 9-11.

Table 6. Estimates of exponential growth models fitted on EXPT2.

Genotype	Parameters	Estimates		
		Height	DBH	Basal
Common	r	0.900	0.97	0.92
	b	-11.24		
	a	7.43		
Auri-PNG	b		-8.51	-9.74
	a		8.67	7.67
Auri-QLD	b		-7.70	-10.80
	a		8.33	8.21
Dive 156	b		-3.84	-6.91
	a		5.55	5.79
Hybr-KX3	b		-5.54	-8.05
	a		6.69	6.43
Mang-PNG	b		-9.72	-11.30
	a		10.00	8.38
Mang-QLD	b		-9.05	-10.75
	a		9.48	8.15
R^2%		90.6	97.1	97.9

R^2% is percentage of variance accounted for.

Figure 9. Height at 6-36 months from equation at Table 6.



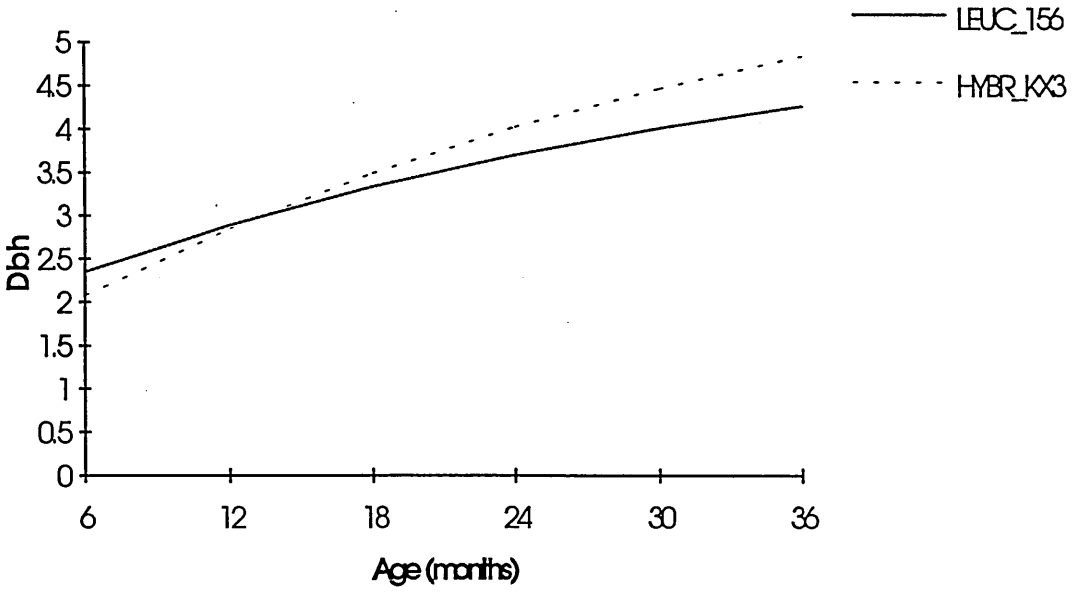


Figure 10. Dbh at 6-36 months for each genotype from equation at Table 6.

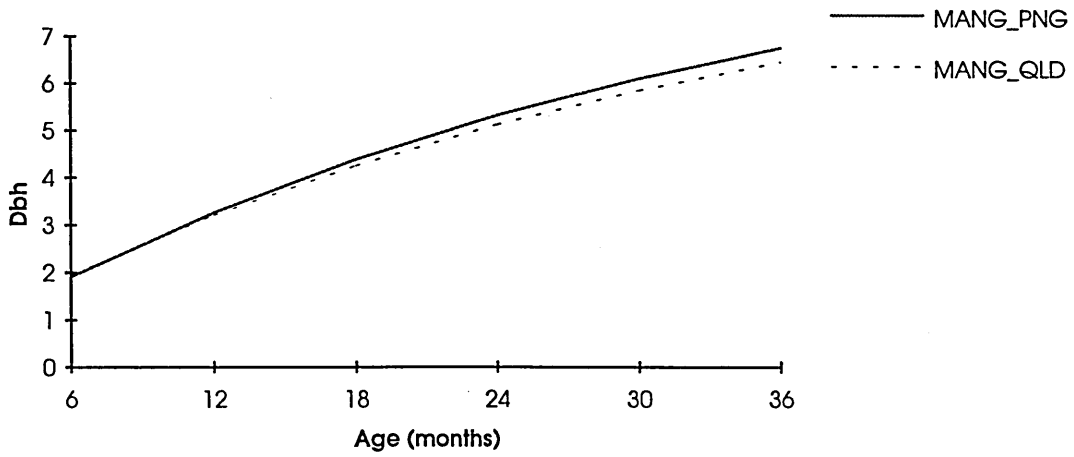


Figure 11. Dbh at 6-36 months from equation at table 6.

Exponential models were also fitted to study the growth behaviour under various cutting methods. The variability in various parameters over cutting techniques can be assessed from table 7.

Table 7. Accumulated analyses of variance when fitting exponential growth models over cutting methods on the EXPT2 data.

Source (change)	d.f.	Mean Squares		
		Height	DBH	Basal
+ Age	2	35.62	6.93	30.72
+ Cutting	2	2.55ns	0.25	0.33
+ Age.Cutting	2	2.24ns	0.14ns	0.14ns
+ Separate non-linear	2	1.70ns	0.08ns	0.09ns
Residual	9	0.91	0.24	0.078

ns - not significant (at 5% level)

The fitted equations to the cutting management treatments are presented are table 8 and the graph in figure 12-13.

Table 8. Estimates of exponential growth curves to the cutting management treatments in EXPT2.

Method	Parameters	Estimates		
		Height	DBH	Basal
Common	r	0.90	0.97	0.92
	b	-11.24	-7.31	-9.62
	a	7.43		
Control	a		7.82	7.21
Pollard	a		7.89	7.34
Pruning	a		8.23	7.67
$R^{2\%}$		74.1	93.6	97.6

Figure 12. Height for cutting managements at 6-36 months from equation at Table 8.

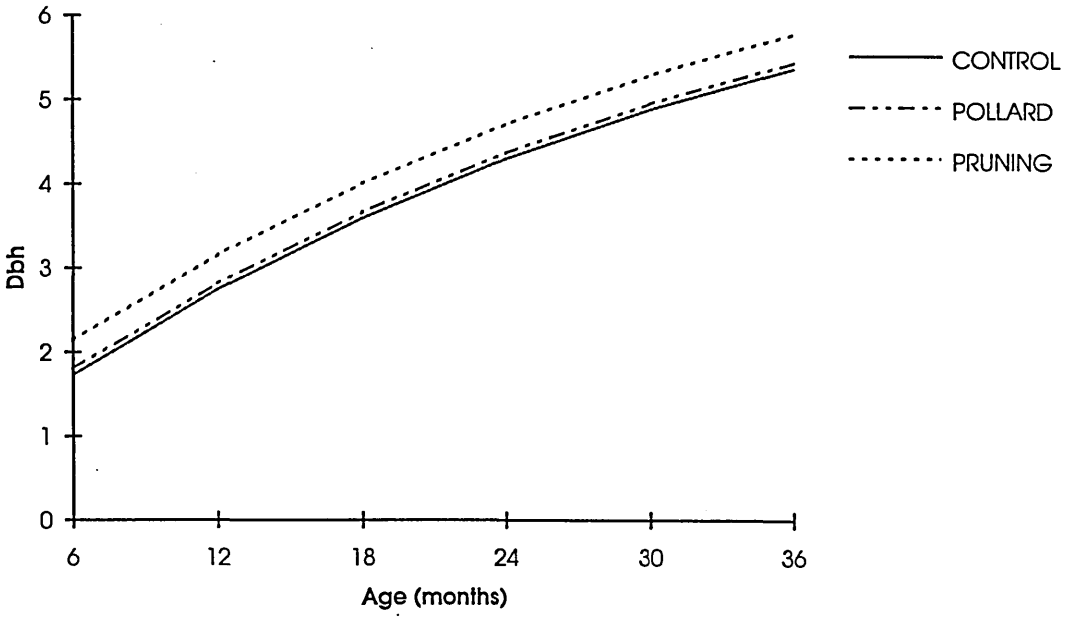


Figure 13. Dbh at 6-36 months for three cutting management treatments from equation at Table 8.

It may be noted from the table 7,8 that all the three methods of cutting result into same growth models for height while for diameters differences in growth is due to different intercepts associated with methods. The exponential models for diameters have explained more than 95% of the total variability.

4.3.3 Logistic Growth Models for Genotypes

The following logistic growth model was fitted on survival percentage.

$$Y = a + \frac{c}{1 + e^{-b(t-M)}}$$

a, b, c and M are the parameters of the model and requires estimation. Parameter M stands for the point of inflection, b a shape parameter, c slope and a + c is asymptote of the variable (ie the value the variables would attain at time infinity or the maximum values attained in this case). This model has been already described in chapter 3 under a different parameterization setup. The analysis of variance and the estimated coefficients obtained at 20th iteration in the numerical computation of the estimates in presented in table 9.

Table 9. Accumulated analysis of variance for Survival percentage.

Source (change)	d.f	Mean Squares
+ Age	3	672.53
+ Genotype	5	55.21
+ Age.Genotype	5	34.28
+ Separate non-linear	10	8.58 ns
Residual	12	8.38

ns - not significant at 5% level.

In order to fit the logistic curve the following GENSTAT commands were given.

```
MODEL  y
TERMS  Age*Genotype
FITCURVE  [curve=logistic]  Age
ADD  Genotype
ADD  Age.Genotype
ADD  [PRINT=m,s,c,a; nonlinear=s]
```

There are significant differences in the intercepts (a) of the genotypes as well as in the slopes (b as indicated by significant interaction Age.Genotype) of the genotypes, the estimates presented in table 10 were obtained.

Table 10. Estimates of common parameters b and M, and genotype dependent parameters c and a.

Genotype	Parameter	Estimates
Common	b	-0.222
	M	18.65
Auri-PNG	c	22.24
	a	79.34
Auri-QLD	c	10.03
	a	90.03
Dive-156	c	20.76
	a	79.63
Hybr KX3	c	21.24
	a	79.35
Mang-PNG	c	31.55
	a	71.17
Mang-QLD	c	20.85
	a	79.57
R^2 %		88.8

The growth behaviour of the genotypes and cutting methods for other experiments can also be estimated using the models described in chapter 3. The few experiments discussed in this chapter have been presented only for illustration.

5 Growth Database - GROWDAT

5.1 Purpose of GROWDAT

The study of plant growth is one of the essential components of research in physiology. The data sets for such research are in the form of time series and differ from the data sets of other plant disciplines which require one time observations on experimental units. Due to this specialized nature of data sets, it was felt necessary to develop a database for storage and retrieval of growth data from a series of experiments, as a part of the thesis work.

GROWDAT has been developed after several consultations with a number of scientists comprising of physiologists, soil scientists, breeders.

5.2 Structure of GROWDAT

The database is user-friendly, easily accessible, menu-driven system. The user need not have previous knowledge or training in computers or database management. The database can be effectively used for storage and retrieval of data for different growth parameters at the replicated level and also at the treatment mean level. The database has utilities for search, add, modify, delete and browse records.

GROWDAT is composed of six data forms and 2 applications. They are, Form A contains Project Description, Form B - Site Description, Form C - Management History, Form D - Botanical Identity, Form E - Factor/Treatment Definition and Form F - Performance Data. The application programs are General Search capability and Data sets for Analysis. The present status of the database creates an ASCII file for transfer to any other statistical package for analysis. Experiments in these forms are linked by a common Experimental Identification Code.

5.3 Future Plan of Work

It is worth exploring a possibility of adding a decision support system to GROWDAT to make it a full fledged database system for storage, retrieval and analysis of growth data as there are no such systems available now exclusively for regression techniques.

The software package known as 'dBase III plus' was used for creating the databases and developing application programs. There are 11 database files, 20 list files for help commands with 120 programs for various functions in GROWDAT. A compiled version of GROWDAT is available for users on request.

6. Conclusion

Statistical modelling is an important tool for quantifying behaviour of plant growth over time in terms of the functions of time. The parameters of the function measure the influence of the factors (alone and in combinations) affecting the dynamics of the plant development. These parameters then facilitate the screening of the plants of desired types. In this study a number of models with linear, polynomial and non-linear forms have been reviewed from statistical fitting and application in plant sciences point of view.

It is observed from that data sets that have been subjected to fitting various non-linear models that the growth behaviour of *Acacia auriculiformis, Acacia mangium* and *Leucaena* trees can be modelled by exponential and logistic curves as in many cases they have explained 90-95% of variation accounted for. The comparisons of the growth behaviour for height, diameter at breast height and survival percentage, were done over genotypes and the cutting methods. The estimates were obtained using GENSTAT5 statistical software. It would be worth exploring in future the application of models, more general in functional form, the treatment of more than one variable jointly to incorporate their association, and various types of error structures in improving the precision of the estimates of the parameters of the growth models.

This thesis also presents the development of GROWDAT, a database for storing, retrieving and manipulating the data generated from a wide class of growth studies. I aim to extend this database to encompass a series of standard statistical analyses and graphical modules in future to help the scientists in analysis and reporting of the research results more efficiently.

7. REFERENCES

Ackoff R L and Sasceni M W (1968); Fundamentals of Operational Research; John Wiley and Sons

Causton D R (1967); Some mathematical properties of growth curves and applications to plant growth analysis. Ph. D Thesis, University of Wales; Aberyswyth.

Causton D R (1969); A Computer program for fitting the Richards Function; Biometrics, 25, 401-9

Causton D R (1970); Growth Function, Growth analysis and Plant Physiology, Forest Commission Research Division, Statistics Section Paper No. 151, Farnham, Surrey

Causton D R (1977); A Biologists mathematics; Edward Arnold, London

Causton D R and Venus, J C (1981); The biometry of Plant Growth. Edward Arnold, London

Checkland P B (1972); Towards a system based methodology for real world problem solving. J. Systems Eng. 3, 2: Reprinted in Systems Behaviour (ed Open Systems Group), Harper & Row, London

Davies O L and Ku J Y (1977); Re-examination of the fitting of the Richards Growth function; Biometrics, 33, 546-547

Hadley P (1978); Growth and Photosynthetic Activity of individual leaves of Plants during growth. Ph.D Thesis, University of Wales

Glasbey C A (1979); Correlated Residuals in non-linear regression applied to Growth data; Appl Statist, 28, 251-259

Gilchrist W G (1984); Statistical Modelling; John Wiley and Sons

Hunt R (1978); Plant Growth Analysis studies in Biology NO. 96; Edward Arnold, London

Hunt R (1982); Plant Growth curves; Edward Arnold, London

Jeffers J N R (1982); Modelling; Chapman and Hall.

Meadows D H (1980); The unavoidable a Priori: In Randers, J (ed) Elements of the System Dynamics Method; The MIT Press, Massachusetts.

Morgan G W (1986); Statistical Models for Biological Growth Curves, Ph.D Thesis, Sheffield City Polytechnic

Nelder J A (1961); The fitting of generalisation of the logistic curve, Biometrics, 17, 89-110

Nelder J A (1962); An alternative form a generalisation of the logistic equation, Biometrics, 18, 614-616

Oliver F R (1964); Methods of estimating the logistic growth functions, Appl Statist 13, 57-66

Oliver F R (1966); Aspects of maximum likelihood estimation of the logistic growth function; J Amer Statist Asst, 61, 697-705

Oliver F R (1982); Notes on the logistic curve for human populations, J R Statist Soc A, 145, 359-363

Patton R A and Krause G F (1972); A parameterisation of the logistic function which admits uncorrelated maximum likelihood estimators; Biometrics, 28, 469-474

Radford P J (1967); Growth Analysis formulae - their use and abuse, Crop Science, 7, 171-5

Randers J (ed) (1980); Elements of the Systems Dynamics Method; The MIT Press, Massachusetts.

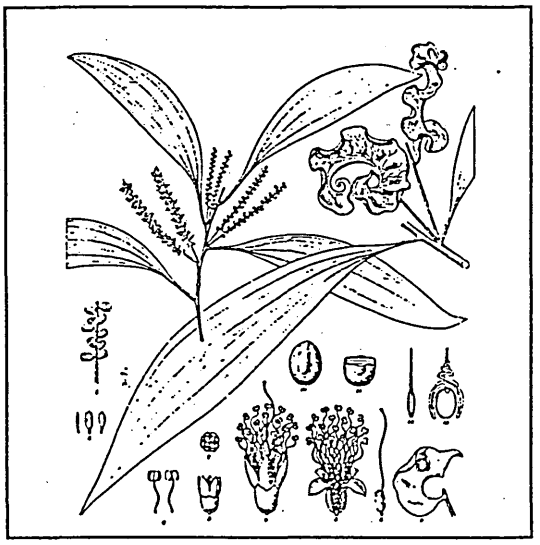
Richards F J (1959); A flexible growth function for emperical use; Journal of Experimental Biology, 10, 290-300

Richards F J (1969); The qualitative analysis of growth, Plant Physiology, Vol VA (Ed Steward FC), 3-76

Solomon, M E (1976); Population Dynamics, 2nd Edition, Studies in Biology No. 18, Edward Arnold, London

Winson C P (1932); The Gompertz curve as a growth curve. Proceedings of the National Academy of Sciences of the United States of America, 18, 1-8

Acacia auriculiformis



Origin

Australia and Oceania, including Papua New Guinea.

Ecology

Mean annual rainfall (mm): 1543 (minimum: 1000, maximum: 3000)

Annual mean minimum temperature (°C): 23 (min: 22)

Annual mean maximum temperature (°C): 32

Altitude range (m): 0 - 700

Dry season adaptability: 4-6 months maximum.

Soils: Can grow on soils ranging from highly acid to alkaline; from shallow clay soils to deep sandy loams.

Light requirement: Strong.

Other site limitations: Low wind tolerance.

Description

Height at maturity: 15-25 meters

Diameter at breast height (1.3 m) at maturity: 50-60 cm

Form: Poor; crooked stem, heavy branching.

Coppicing ability: Poor-Fair

Growth: 15-18 m in height, 15-20 cm in diameter in 10-12 year rotation.

Other: Fixes nitrogen, produces high levels of nitrogen even on poor soils.

Primary advantages

Good pulp production on highly infertile sites with pH as low as 3.0 or on soils as shallow as 20 cm; also can shade out imperata grass (*Imperata cylindrica*).

Primary disadvantages

Has strong allelopathic effects that limit tree-crop interactions.

Products and Yields

Wood products: Poles, pulp, timber; 12-15 cubic meters/ha/year on 10-12 year rotation, higher or lower yields depending on rainfall and soil type.

Fuelwood: 4600-4800 kcal/kg, yields of 16 tons/ha/year; branches and leaves are a good source of small diameter fuelwood, yielding 4-6 tons/ha/year.

Fodder: Unpalatable for livestock.

Other: Shade tree, tannin and gums.

Propagation

Planting seedlings is the best method but, it is possible to direct seed.

Seed treatment

Some scarification of the seed is needed, hot water scarification is best.

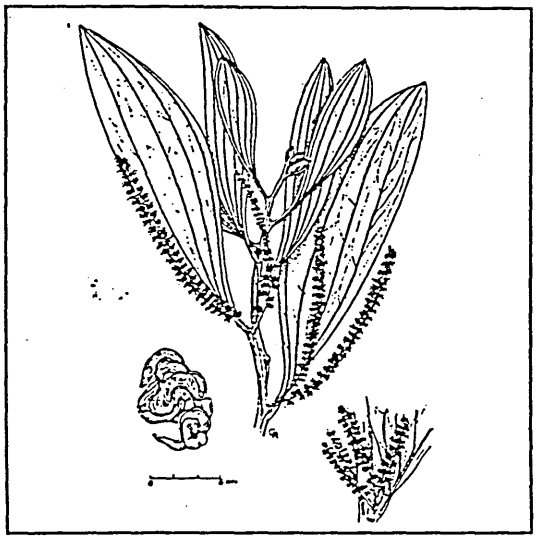
Management

Pruning, coppicing, pollarding. For fuelwood and pulp, usually planted at 2 x 2 m up to 4 x 4 m. Closer spacing is better for fuelwood. For better coppicing, cut stumps at 1 m above ground. Removing lower branches of young plants may improve stem form.

Agroforestry Uses

Limited by the allelopathic effects of prolific leaf litter. Some uses include hedgerows on steep denuded lands. Used as firebreak species in orchards, shades out imperata grass, thus reducing fire danger. Not recommended for growing close to food crops.

Acacia mangium



Origin

Australia and Oceania, including Papua New Guinea

Ecology

Mean annual rainfall (mm): 1381 (min: 1000, max: 4000)
Annual mean minimum temperature (°C): 20.8 (min: 17.5)
Annual mean maximum temperature (°C): 30.4 (max: 31)
Altitude range (m): 0 to 720
Dry season adaptability: Poor
Soils: pH from 4.5-8.0, grows well on red-yellow podsols, even if heavily eroded. Can tolerate some waterlogging.
Light requirement: Strong, as it is a pioneer species.
Other site limitations: Performs poorly with less than 1200 mm annual rainfall; does not tolerate wind.

Description

Height at maturity: 25-30 meters
Diameter at breast height (1.3 m) at maturity: 40-60 cm
Form: Good, self pruning, straight bole without knots, especially when grown in plantation.
Coppicing ability: Only in young stems, poor in old trees.
Growth: In a 13-year plantation, can reach 23-25 m height,

27-30 cm diameter at breast height (1.3 m).
Other: Fixes nitrogen.

Primary advantages
Provides timber and other wood products; can quickly suppress imperata grass (*Imperata cylindrica*) on degraded acid soils. At present, few problems with pests.

Primary disadvantages
Some damage has been reported in young stands due to pinhole borers; some damage in nursery due to mildew and molds; wood has a high degree of 'spring' in milling test, a potential defect; heart rot can be a problem in older stands.

Products and Yields
Wood products: Timber, pulp, plywood, particle board. Yields range from 14 cubic meters/ha/year at 4 years to 44 cubic meters/ha/year at 10 years. Its timber has nice, close grain.
Fuelwood: 4,800-4,900 kcal/kg, produces high quality charcoal.
Fodder: Generally considered a poor fodder tree.
Other: Shade and ornament along roadsides and in gardens.

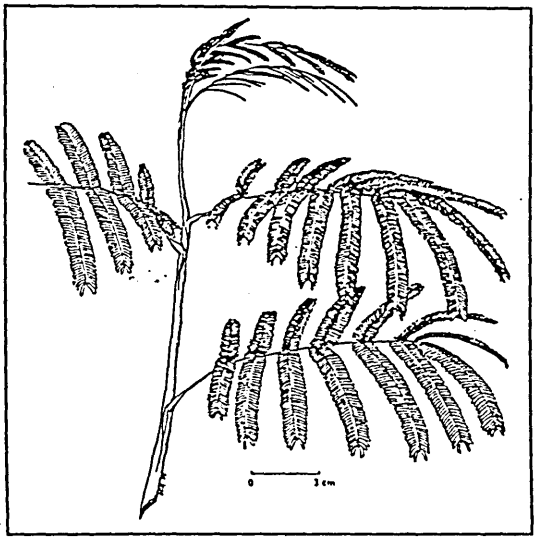
Propagation
By seedling. No fertilizer needed unless soil contains no phosphorus (P).

Seed Treatment
Boil water and pour it over the seed, 1 part seed to 10 parts water. After 30 seconds to 1 minute, remove seeds and place in tap water; then soak overnight, remove, and dry.

Management
Commonly planted at 3 m x 3 m spacing. For timber, prune every six months, up to age 2 years, and thin stands at ages 2, 6, and 10 years. To avoid heart rot, harvest at 5 - 7 years.

Agroforestry Uses
Used in taungya systems. Closes canopy quickly, however, and must be planted in wide spacing (greater than 3 m x 3 m) to allow more than one year of annual inter crop.

Leucaena diversifolia



Origin
Central America

Ecology
Mean annual rainfall (mm): 600 to 2800
Annual mean temperature range (°C): 18 to 30
Altitude range (m): 500 to 2000
Dry season adaptability: Not good; no long periods of drought.
Soils: Wetter soils than *Leucaena leucocephala*.
Light requirement: Strong.

Description
Height at maturity: 20 meters
Diameter at breast height (1.3 m) at maturity: 20 cm
Form: Good, fairly straight bole for several meters is possible.
Coppicing ability: Excellent.
Growth: Fast.
Other: Fixes nitrogen.

Primary advantages
Fast growing for light poles and fuel; grows at higher altitudes

Primary disadvantages

Not as useful at low altitudes; fodder quality not as high as *L. leucocephala*, although lower mimosine content.

Products and Yields

Wood products: Poles, timber, and fuel. Poles are of poor quality.

Fuelwood: Wood density of 0.4-0.5, provides 4,500-4,600 kcal/kg.

Fodder: Can be mixed as feed ingredient.

Propagation

By planted seedlings. New hybrids with *L. leucocephala* have been developed in Taiwan and Hawaii, U.S.A.

Seed Treatment

Soak in hot water for 2-3 minutes, then in room temperature water for 2-3 days. Sow soon after treating.

Management

Coppicing, lopping, pollarding. Rotations of 6 months for small fuelwood to 10 years for poles. Fodder can be harvested after 2-4 months in hedgerow systems.

Agroforestry Uses

Hedgerows, alleycropping, intercropping. Hedgerows produce valuable green manure, fodder, and small fuelwood and increase soil nitrogen. Commonly intercropped with cassava, papaya, and sweet potato.

PROGRAM DESCRIPTION

Its six data entry components (forms) include:

A - Project Description	D - Botanical Identity
B - Site Description	E - Factor/Treatment Definition
C - Management History	F - Performance Data

The form options are search, add, modify, delete, and browse. These options operate on records in a specific form.

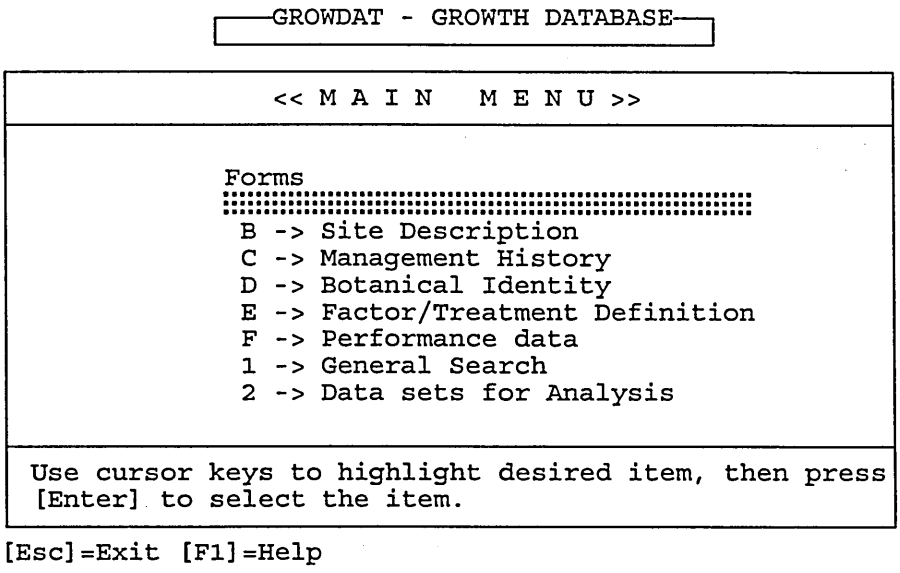
The program also includes two applications:

1 - General Search
2 - Data sets for Analysis

The general search retrieves and displays data from all six data forms. Growth performance records selected with this option can be saved and used for step-wise searches or for creating data sets for analysis. The data sets and analysis creates data sets from selected variables.

MAIN MENU

GROWDAT is composed of 6 data forms and 2 applications. The main menu lists the options.



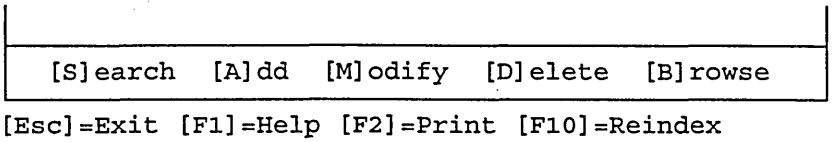
Screen 1. Main Menu

To select an option use [↓] and [↑] keys to move the highlight bar to the option you desire and press [Enter]. You may also select an option by pressing the highlighted letter or number in front of the option.

NOTE: This type of menu is called highlight bar menu. The other type of menu is the highlighted key menu described in the following section.

DATA FORM OPTIONS

After selecting one of the data forms you can choose from the options displayed at the bottom portion of the screen. An option is valid when its key label is displayed in bright type and is selected by pressing that key.



Screen 2. Data Form Options

[S] - displays search fields where you can enter search criteria to retrieve and display existing data. All records are selected if the search fields are left blank. Searching is initiated by pressing <Ctrl-W>. The records found during a search may be modified by pressing [M], delete by pressing [D], or printed by pressing [F2].

[A] - displays the data entry fields. Data entered is saved by pressing *<Ctrl-W>*. Pressing [Esc] abandons the data entry.

NOTE: To maintain the linkages between the forms certain fields require an entry. These are the ID fields. Examples of ID fields are the Project ID, Site ID, and Experiment ID.

[M] - automatically puts you in the Search mode since you first need to find records to modify. Pressing [M] after a successful search displays the data entry fields containing the existing data for modification.

[D] - automatically puts you in the Search mode. Pressing [D] after searching, causes a message to appear asking you to confirm the deletion of the displayed record. Deleted data cannot be recovered and requires re-entering the data if it is needed again.

[B] - displays the contents of the file containing the data records for a form. Data is displayed as they are stored in the file. Each column represents a field and each row is a record. The field names appear at the top of each column.

[Esc] - exists the form and return to the previous screen.

[F1] - displays a help screen which briefly describes the options.

[F2] - outputs the data of the current record to a printer. This key becomes active after a successful search.

[F10] - reindexes a form's corresponding database file. Index files facilitates the retrieval of data records and may occasionally be corrupted. Use this option to recreate the index files if difficulties are encountered while searching.

NOTE: *<Ctrl-W>* means to hold down the *[Ctrl]* key and press *[W]*.

SUBMENUS

Selecting some of the menu options displays submenus. Options on the submenus are selected like the options on the main menu. You can move the highlight bar to the desired option and then press [Enter] or press the key appearing in bright type at the beginning of the option.

HELP WINDOWS

During data entry you can press [F1] to display a brief description or example of acceptable data. The help window is removed by pressing any key. Help windows are also available for the options that appear in the lower portion of the screen. Whenever the 'F1' appears in bright type a help window is available by pressing [F1].

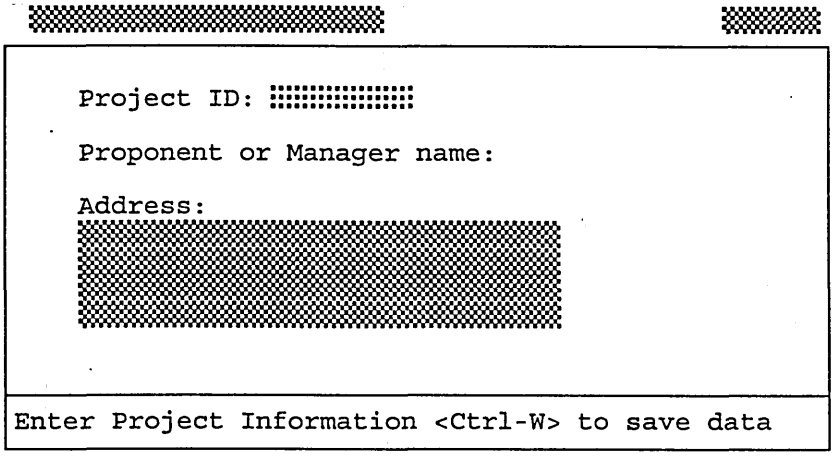
DATA FORMS

The data forms accept and display project, site, experiment, species/genotype, and performance information. The labels for data items displayed in upper case letter signify minimum data set items and should be provided. Data item labels displayed in upper and lower letters are not part of the minimum data set and are preferred data.

The data forms are shown in the Add mode to illustrate the size of the input fields.

Project description

Form A contains the information that identifies the project responsible for experiments.

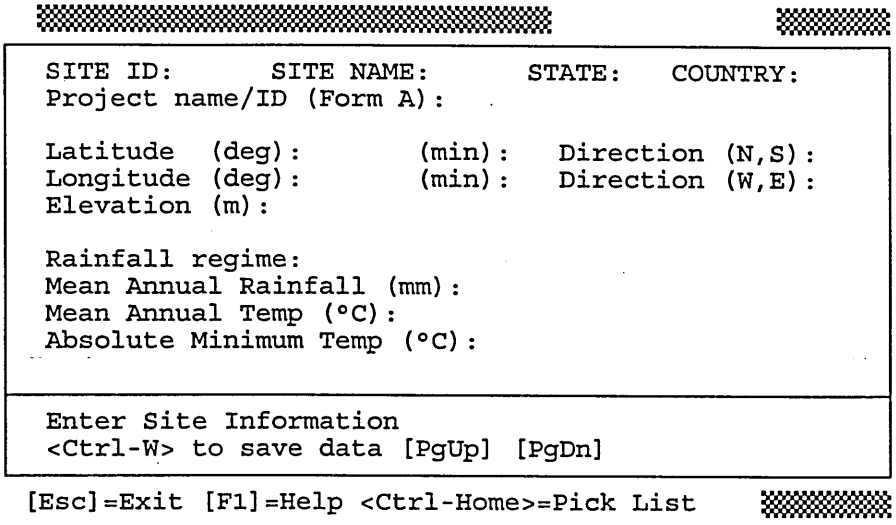


Screen 4. Form A. Project Description

Project ID is the key field in this form. It is used in Form B to link a site with a project.

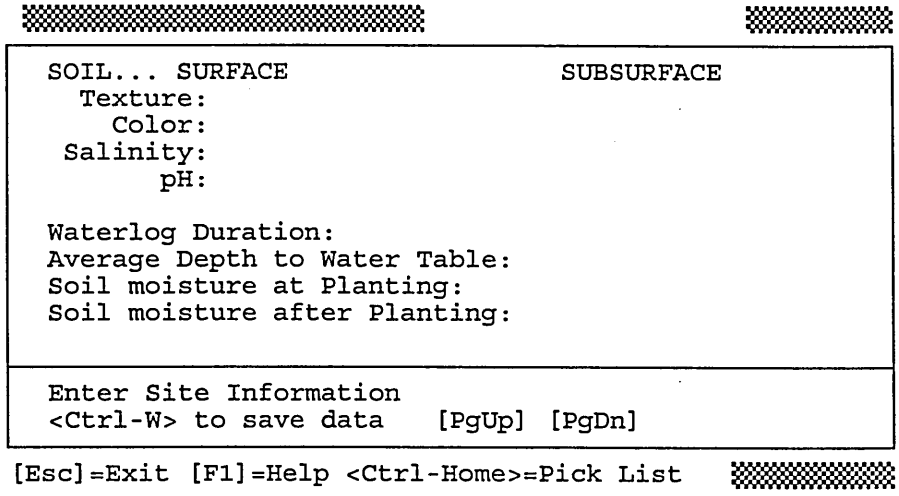
Site description

Form B contains the information describing the site of the experiment. It is composed of two pages (screens). They contain the geographic position, climate, topography, and soil characteristics of the site. Planting information of the experiment is also entered. A site is linked with a project by entering the appropriate Project ID.



Screen 5. Form B. Trial Site Description, Page 1

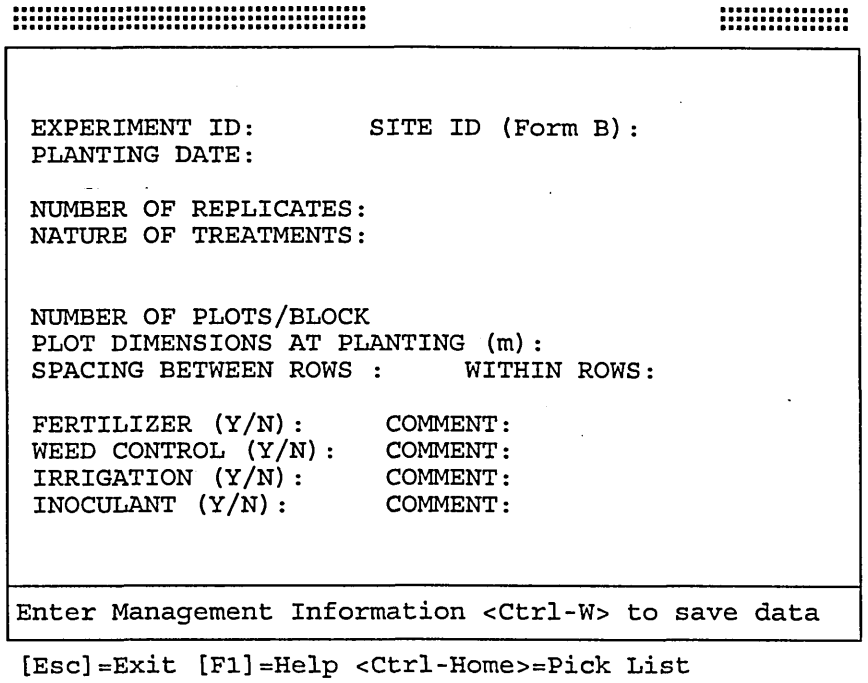
SITE ID is the key field in this form. It is used in Form C to link an experiment with a site.



Screen 6. Form B. Trial Site Description, Page 2

Management History

Form C contains information on the experiment. It is linked to its site by entering the appropriate SITE ID.

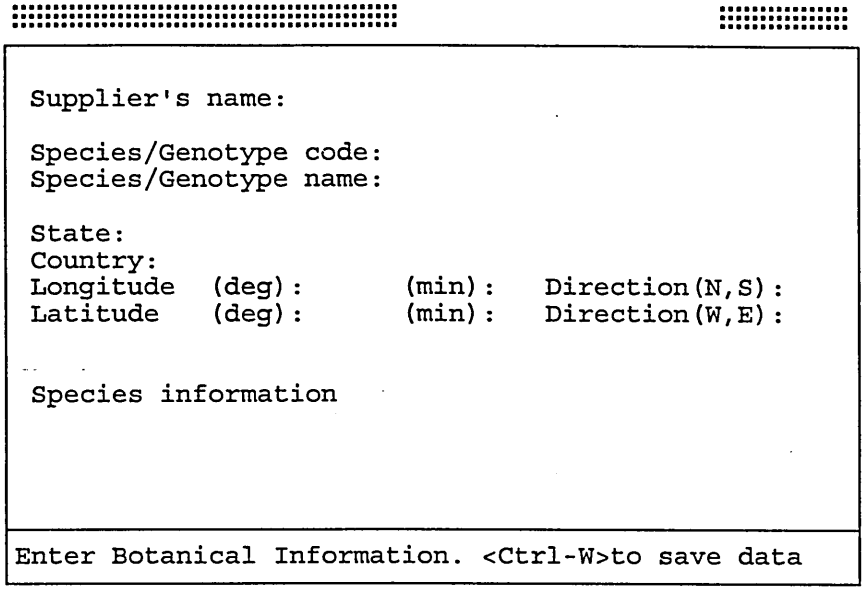


Screen 7. Form C. Management History

EXPERIMENT ID is the key field in this form and many other forms. It is an important ID in the database in that it serves to link this form with the factors/levels, treatment, and performance data.

Botanical Identity

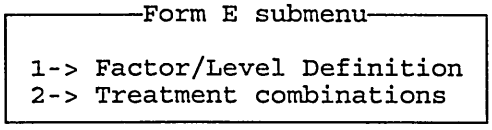
Form D contains information on the species/genotype used in the experiment. A record is entered for each genotype or species. Seed origin is the key in this form.



Screen 8. Form D. Botanical Identity

Factor/Treatment Definition

Form E is broken down into sub-forms. Form E-1 is where you enter experiment factor and level information. In Form E-2 you define the treatment combinations. Each subform is selected by using the submenu. The factors and levels have to be defined (in E-1) for the experiment before its treatments can be entered (in E-2).



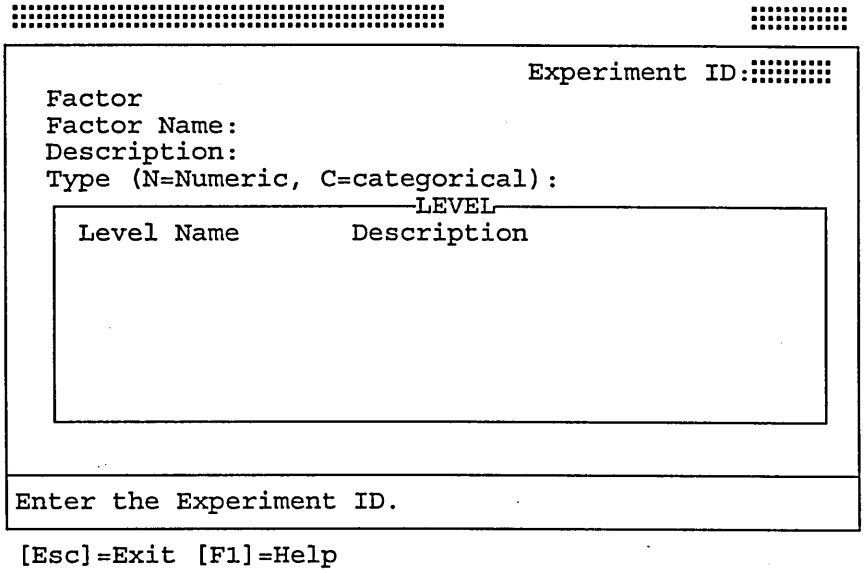
Screen 9. Form E Submenu

Factor/Level definition

Form E-1 contains information on the factors and levels of each factor used in the experiment. This information is entered by experiment.

First enter an Experiment ID (6 character code representing the experiment), then you enter the factor ID (A,B and C), name, description, and type. The factor type determines if levels are required in the definition. Unlike categorical factors, numeric factors donot have predefined levels.

NOTE: The factor type (numeric or categorical) cannot be changed once treatments are defined.

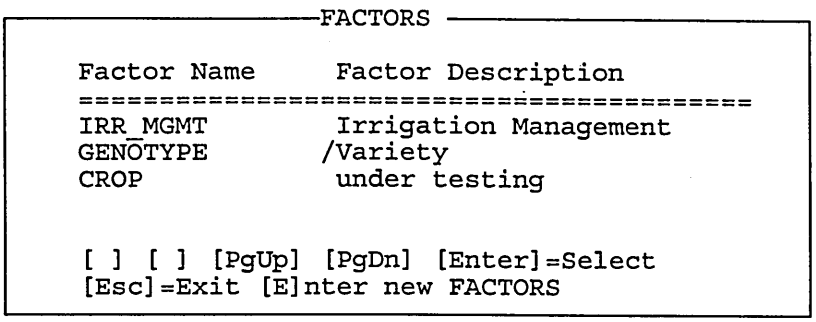


Screen 10. Form E-1. Factor/Level Definition

The Experiment ID, the key field in this form, is entered first in the upper right corner of the form.

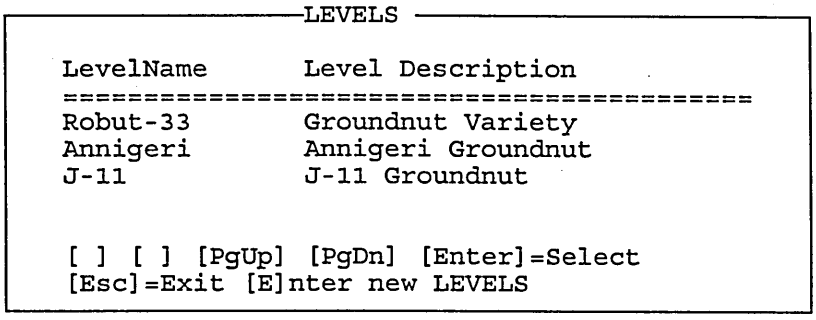
NOTE: The factors and levels previously defined in the database should be used if possible to avoid extraneous factor and level definitions, that is, different factor and level names that actually mean the same thing. Using the same factors and levels as other experiments will allow you to later retrieve experiments based on their factors and levels.

The factor name can be entered directly or selected from the Factors Selection Window (press <Ctrl-Home>). It lists the factors contained in the database. If your experiment factor is not listed you can add that factor to the list.



Screen 11. Factors Selection Window

The levels of a factor can be entered directly or selected from the Levels Selection Window (press <Ctrl-Home>). It lists the levels in the database for the selected factor. If your experiment uses levels not included in the list you can add those levels to the list.



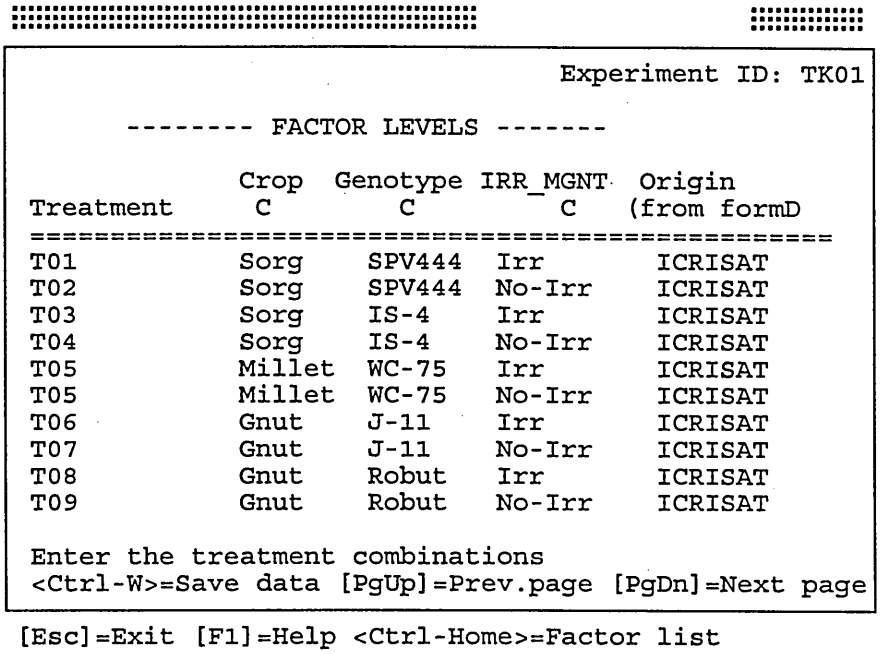
Screen 12. Levels Selection Window

NOTE: The Factor Selection Window and the Levels Selection Window will appear automatically if there are no factors or levels defined in the database.

New factor and level names can be added by pressing [E]. You need to enter the factor or level name and its description. The name can be a maximum of eight characters and should contain only letters, digits, or the underline character '_'. It must begin with a letter.

Treatment combinations

Information entered in Form E-2 is entered by experiment. First enter an Experiment ID (to link this form with Form C), next enter the treatment names and select the factor level combinations for the treatments. If treatments already exist for the experiment, you have the option to add more treatments and modify the treatments that already exist. The experiment treatments are linked to Form D by entering the seed origin.



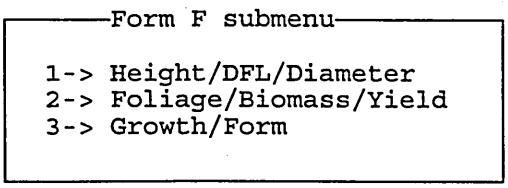
```
                                        Experiment ID: TK01

       -------- FACTOR LEVELS -------

                Crop   Genotype IRR_MGNT  Origin
Treatment        C         C         C    (from formD
================================================
T01             Sorg     SPV444   Irr        ICRISAT
T02             Sorg     SPV444   No-Irr     ICRISAT
T03             Sorg     IS-4     Irr        ICRISAT
T04             Sorg     IS-4     No-Irr     ICRISAT
T05             Millet   WC-75    Irr        ICRISAT
T05             Millet   WC-75    No-Irr     ICRISAT
T06             Gnut     J-11     Irr        ICRISAT
T07             Gnut     J-11     No-Irr     ICRISAT
T08             Gnut     Robut    Irr        ICRISAT
T09             Gnut     Robut    No-Irr     ICRISAT

Enter the treatment combinations
<Ctrl-W>=Save data [PgUp]=Prev.page [PgDn]=Next page
[Esc]=Exit [F1]=Help <Ctrl-Home>=Factor list
```

Screen 13. Form E-2. Treatment Combinations

The treatment name should contain only letters, digits, or the underline character. The levels for each factor can be entered directly or by using <Ctrl-Home> to display the list of levels (only if the factor is categorical) and then selecting from the list.

Performance Data

Form F is divided into three subforms. A subform is selected using the Form F submenu. Move the highlight bar to the desired subform and press [Enter]. You can also press the number corresponding to the desired subform.



Screen 14. Form F Submenu

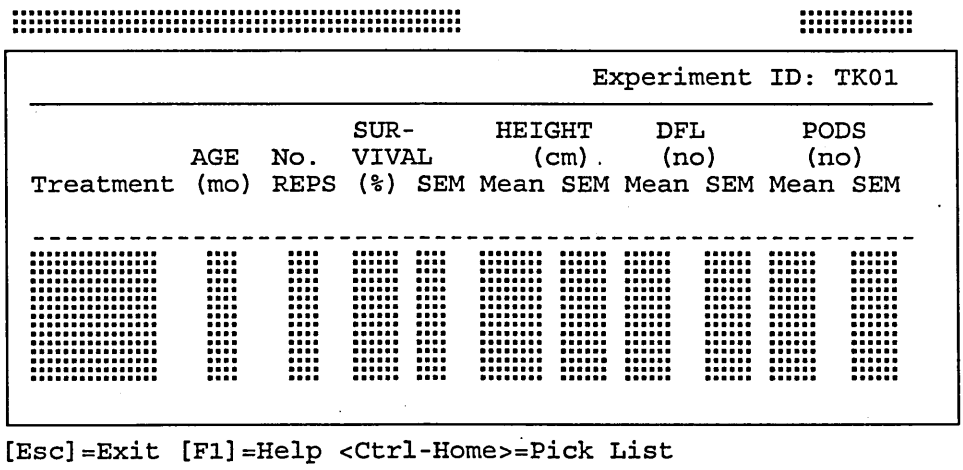
Form F information is entered by experiment. First enter an Experiment ID to link this form with Form E. For each line, enter the treatment (or use <Ctrl-Home> to select from the pick list) and enter the measurements.

In the ADD mode treatment/age combinations cannot be entered. In the MODIFY mode, the treatment/age combination can not be changed to one that already exists. In addition, in the MODIFY mode, leaving either the treatment or age blank is not allowed.

NOTE: In the ADD mode, if you do not enter a treatment name and age, the data will not be saved.

Height/DFL/Diameter

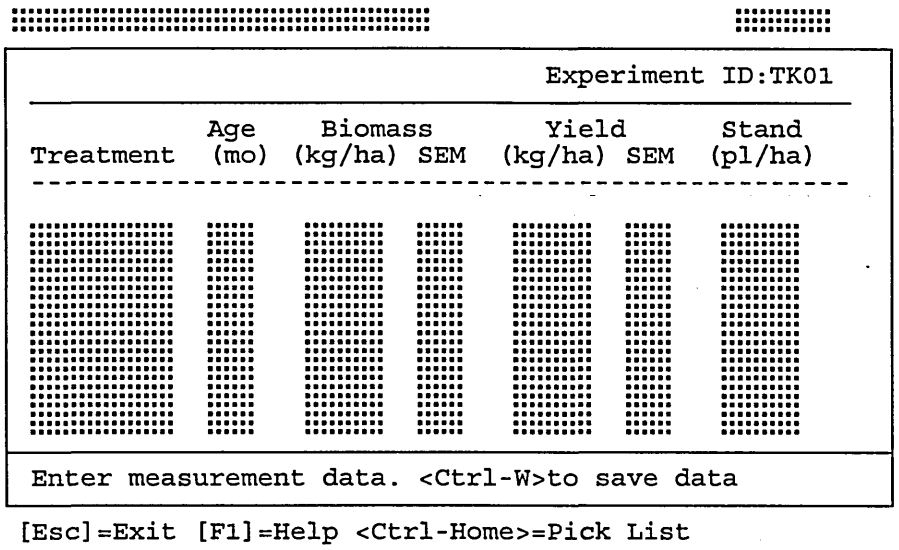
This subform also includes the age, the number of replications used to calculate the treatment mean and survival data. Also in this form is an extra diameter measurement which includes the height or days to flowers and diameter measurements. All measurements in this subform should include the standard error of the mean (SEM).



Screen 15. Form F-1 Growth Performance Data

Foliage/Biomass/Yield Measurements

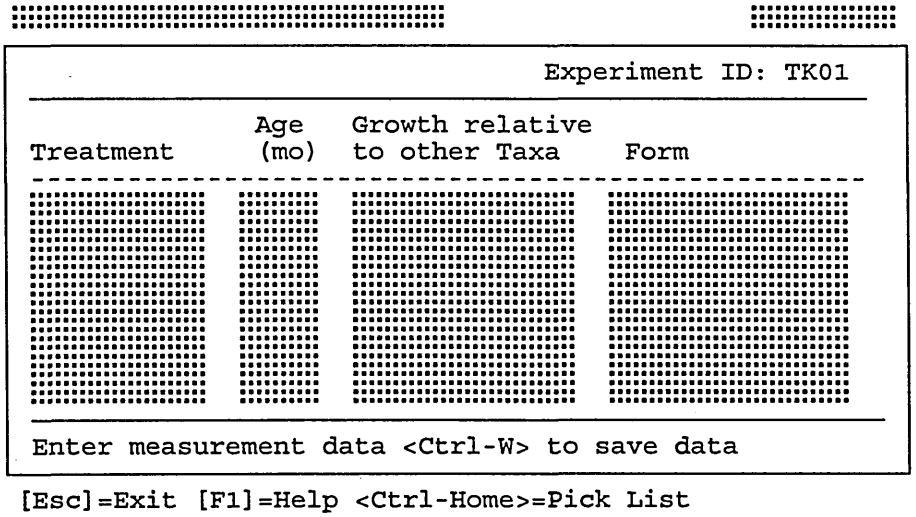
This subform also includes the stand density.



Screen 16. Form F-2. Performance data

Growth/Form measurements

This subform contains the growth and tree form measurements.



Screen 17. Form F-3. Performance data (Growth form)

SEARCHING BY FORMS

This option allows you to retrieve selected record(s) within a single form.

This prototype implements searching by forms in Forms A,B,C,D,F-1,F-2and F-3 only. The records in these forms can be retrieved and displayed on the screen according to a search criteria entered by the user. To search for data, press [S] after selected a form in the main menu. The search fields will appear on the form. Performance data can be searched by experiment ID, seed origin, species/crop code, and/or age range.

To search for all records in the database leave all search fields blank. Press <Ctrl-W> to begin the search or [Esc] to exit to the main menu.

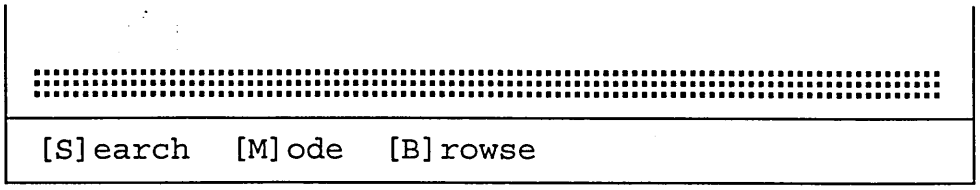
After a search is performed the program will display the records found one at a time for Forms A, B, C and D. You can modify or delete the record being displayed on the screen.

Performance records are displayed eleven records at a time for the subform (F-1, F-2, or F-3) in which the search was done. The records can be modified but cannot be deleted.

GENERAL SEARCH

General search (option 1 in the main menu) is used to retrieve linked records from the database. The search criteria entered are used to select records from the site description (Form B), management history (Form C), botanical identity (Form D), factor/level description (Forms E-1 and E-2) and performance data (Form F) files. The default growth performance data file searched is SUMF.DBF which contains all the data entered using the add option in Form F: Plant/Tree performance Data. Other performance files, saved from previous searches, can be used instead of SUMF.DBF. This allows for step-wise searching.

The operational options (keys) in general search are [S] to display the search fields, [M] to select the search mode, [B] to browse the current file, [F3] to change the current file, and [Esc] to exit and return to the Main Menu. The current file status line displays the filename, the number of records, the number of experiments, and the date of the file.



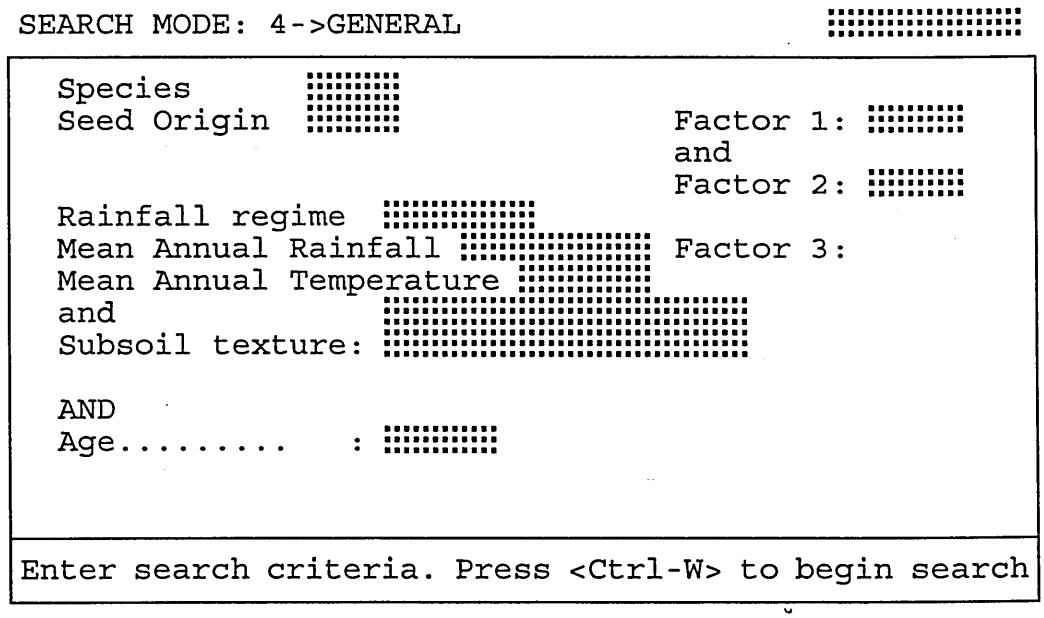
Screen 18. General Search Menu

Search

Pressing [S] displays the search fields. You enter the search criteria into the fields and then press <Ctrl-W> to begin the search procedure.

The search fields are listed in Table 1. Table 1 also indicates the availability of the search fields in the various search modes. The search mode determines the search fields to be displayed. The search fields are grouped into crops/species, climate, soils, age, and factors/levels fields.

The Crops, Species, Seed origin, Rainfall regime, Climate code, Subsoil texture, Factor 1, Factor 2, Factor 3 and Levels of Factor 1 search fields will accept only valid entries. The valid entries are determined by the records in the database.

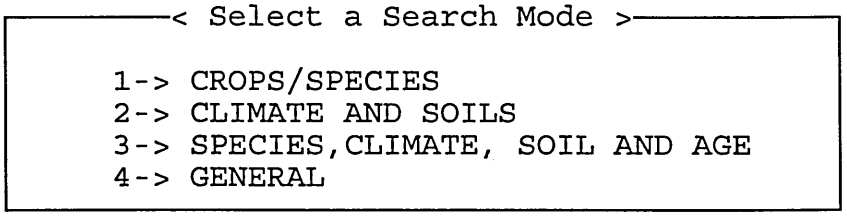


Screen 19. General Search Screen (Mode 4)

NOTE: The Levels of Factor 1 input fields are visible only when a valid entry is placed in the Factor 1 input field. A valid entry means a factor name which exists in the database and has levels defined.

Mode

Pressing [M] displays the Search Option window. Use the arrow keys, [ ] or [ ], to highlight the desired option and then press [Enter] to select. Press [Esc] to exit without selecting a mode.



Screen 20. Mode Selection Window

You have the option of selecting one of the four search modes. Each mode displays a different set of search fields. Table 2 shows which search fields are available for each search mode. For example the species group of fields are available in modes 1,3 and 4 but not in mode 2.

Browse

The browse option allows the user to look at the records contained in the current file without performing a search. The data is presented as they are actually stored in the file. Use this option to quickly see the records in the current file. This could help you select your search criteria.

Change File

The current file is displayed in the file status line (see screen 19). It lists the filename, number of records, number of experiments, and the date of the file. This file can be changed by pressing [F3].

The default current file is SUMF.DBF, the file containing all performance data records. When searches are performed, the records retrieved from SUMF.DBF can be saved to another file with a file name you specify. The saved file can then be used as the current file to perform additional searches.

Exit

The [Esc] is used to exit from the general query module and return to the Summary Database Main Menu.

SEARCH EXAMPLES

These examples show some of the features of the General Search option. The retrieval and display of the records are accomplished by entering search criteria in the fields displayed and then pressing <Ctrl-W> when you want the searching to begin.

Search Mode 1: Species

This mode is used when the species, provenance and/or seed origin are the only search criteria that need to be satisfied.

Example 1. This example shows the search procedure used to retrieve records for a specific species.

Display the species pick list by pressing <Ctrl-Home> while the cursor is in the Species field. Select a species 'ACACAURI' from the list by highlighting an item using the arrow keys and then pressing [Enter].

Example 2. This example shows a search procedure to limit the search to specific provenances within a given species or specific genotypes within a given crop.

Repeat example 1 (using 'ACACAURI' as the species), but instead of pressing <Ctrl-W> press [Enter] to move the cursor to the Genotype field. Press <Ctrl-Home> to display the provenance/genotype pick list. Select 'AURI_PNG' as the genotype/provenance by highlighting it and pressing [Enter].

NOTE: Pressing <Ctrl-Home> displays a selection list for the Species, Provenance/Genotype, and seed origins. The items shown on the genotype list depends on the entry made in the Species field. If the Species field is blank, all genotypes/provenances will be displayed on the list. If a species has been entered only the provenances/genotypes for that species will be shown on the list. The same relation exists between the seed origins and species/genotypes.

Search Mode 2: Climate and Soils

This mode is used for retrieving data based on the characteristics of the site where the experiment was conducted.

Example 1. This example illustrated searching by specifying a range search field.

Enter 1000 and 2000 in the two Mean Annual Rainfall search fields.

Example 2. This example shows the use of a climate and a soil search field for retrieval.

Repeat example 1 but also select 'Clay(Light clay, Heavy Clay)' from the pick list in the Subsoil Texture search field.

Example 3. This example shows the use of a logical operator.

Repeat example 2 but change the 'AND' to an 'OR'.

NOTE: The 'AND' displayed between the climate and soil search fields is a logical operator. Its valid entries are 'AND' and 'OR'. The 'AND' operator causes the retrieval of records that match the climate 'AND' the soil search criteria. The 'OR' operator causes the retrieval of records that match the climate criteria 'OR' that match the soil criteria.

Search Mode 3: Species, Climate, Soils and Age

This mode combines modes 1 and 2 and adds the capability of selecting performance records based on the time of measurements.

Example 1. This example shows the use of the Age search field.

Enter '12' and '18' in the two Age search fields.

Example 2. This example shows the use of the Age search field with the Species search fields.

Enter a species in the Species field and repeat example 1.

Example 3. This example shows that when a Species is entered (with no other search criteria), the logical operator above the Age search fields has no effect.

Repeat example 2 but replace the 'AND' above the Age fields with an 'OR'.

NOTE: The species fields (Species, Provenance/Genotype, and seed origin), when filled, are automatically 'AND'-ed with other search fields. That is, the records that will be retrieved will satisfy the species fields' criteria.

Search Mode 4: General

This mode is similar to mode 3 but includes the capability of specifying the factors (upto three) and levels of one factor (Factor 1) in the search criteria. This mode is useful in selecting performance records from experiments with common factors/levels in their treatment designs.

Example 1. This example shows the search procedure to find records with a specify factor in the treatment design.

Display the factor pick list by moving the cursor to the Factor 1 search field and then pressing <Ctrl-Home>. Select 'SPECIES' from the list.

Example 2. This example shows the search procedure to find record with two specific factors in the treatment design.

Select 'SPECIES' for Factor 1 and 'SPACING' as factor 2.

Example 3. This example shows the search procedure to find records with at least one of the two given factors in the treatment design.

Use 'GENOTYPE' for Factor 1, replace the 'AND' between Factor 1 and factor 2 with 'OR', and use 'SPACING' for Factor 2.

NOTE: <Ctrl-Home> displays a list of ALL existing factors in the database. The existence of a factor in the list does not guarantee the existence of performance records with that factor.

NOTE: The default logical operator between the factors is 'AND'. The 'OR' operator can also be used. Note that 'AND' takes precedence over 'OR'.

Example 4. This example shows the search procedure used to retrieve records with specific factor in the treatment design and a specific level of that factor in the treatment combination.

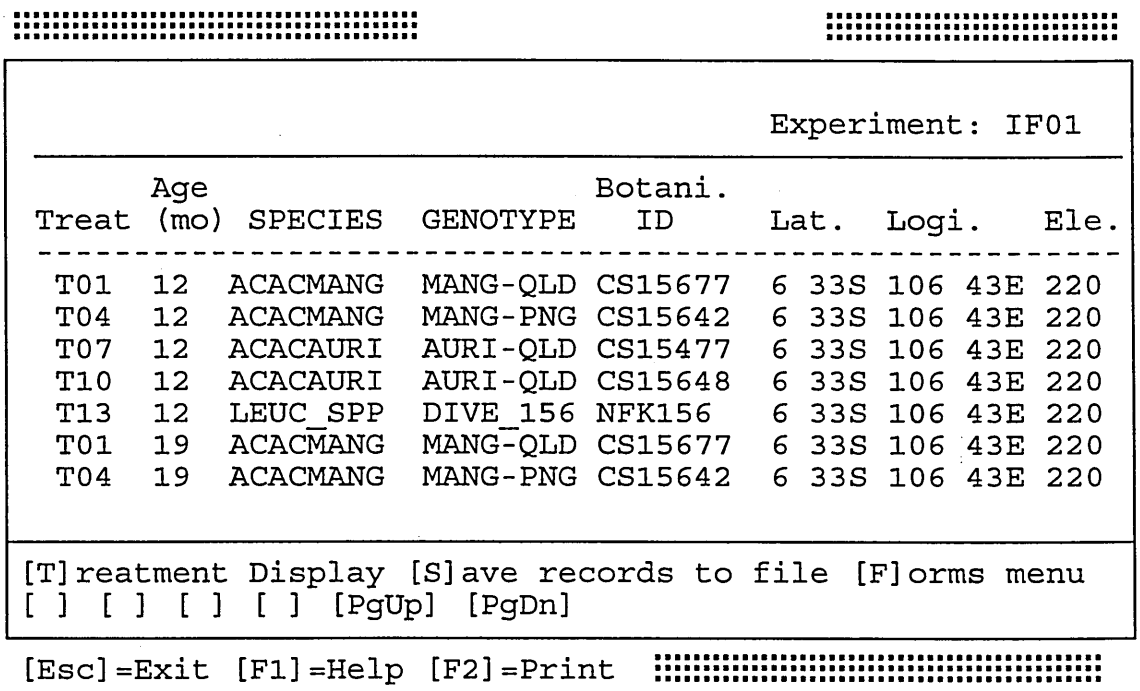
Select 'SPECIES' for Factor 1. Move the cursor to the first line of the Levels of Factor 1 search list and press <Ctrl-Home>. Select 'EUCACAMA' from the list.

NOTE: You may enter as many as eight levels for the factor specified in Factor 1.

Summary display

After a search is performed the program will display the first screen (page) of a two-screen summary of the records founds. The summary screens are shown in Screen 20 and 21.

The operational options (keys) in these screens are [T] to display the treatment design of the highlighted treatment name, [F] to select a specific form to view, [S] to save the performance records retrieved in a disk file, the arrow keys to move the screen, the [PgUp] and [PgDn] keys to move the display up and down, and [Esc] to return to the search screen.



Experiment: IF01

Treat	Age (mo)	SPECIES	GENOTYPE	Botani. ID	Lat.	Logi.	Ele.
T01	12	ACACMANG	MANG-QLD	CS15677	6 33S	106 43E	220
T04	12	ACACMANG	MANG-PNG	CS15642	6 33S	106 43E	220
T07	12	ACACAURI	AURI-QLD	CS15477	6 33S	106 43E	220
T10	12	ACACAURI	AURI-QLD	CS15648	6 33S	106 43E	220
T13	12	LEUC_SPP	DIVE_156	NFK156	6 33S	106 43E	220
T01	19	ACACMANG	MANG-QLD	CS15677	6 33S	106 43E	220
T04	19	ACACMANG	MANG-PNG	CS15642	6 33S	106 43E	220

[T]reatment Display [S]ave records to file [F]orms menu
[ ] [ ] [ ] [ ] [PgUp] [PgDn]

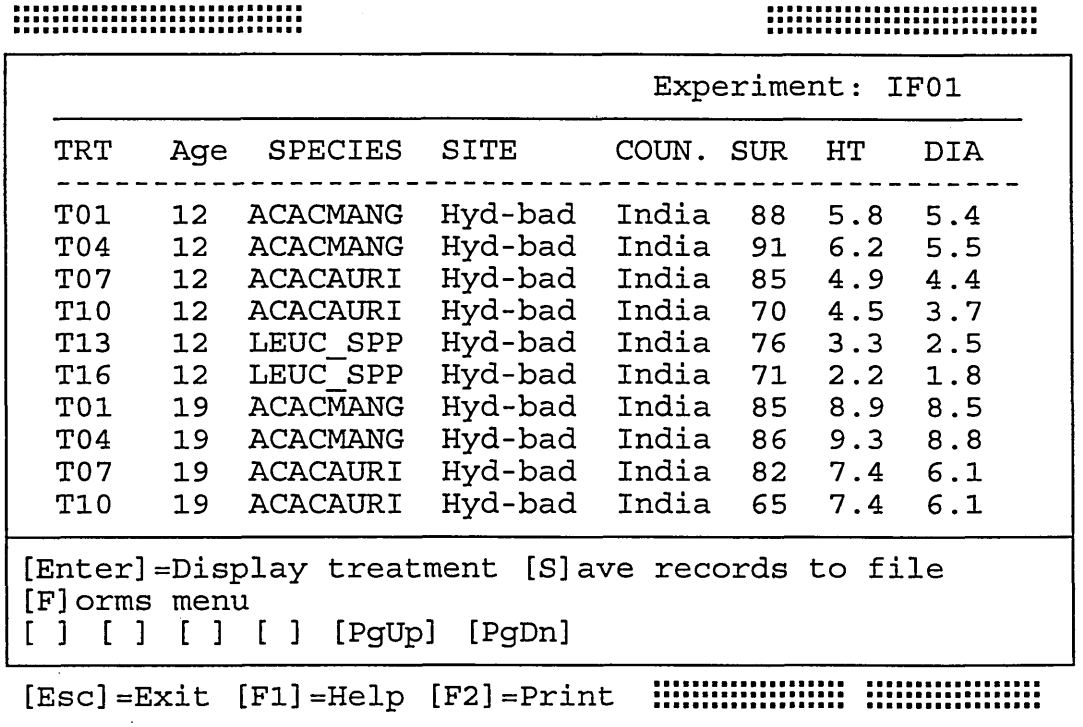
[Esc]=Exit [F1]=Help [F2]=Print

Screen 21. Summary Display, Page 1

Treatment display

You can highlight a treatment (using the up and down cursor key) and press [T] to displaythe experiment factors and treatment combinations for that treatment.

NOTE: The TREATMENT column of the screen lists the treatment names of the performance records. A highlight bar is used to mark a specific line on the display. The highlighted treatment serves two purposes, it determines the experiment ID displayed at the upper right corner and the treatment combination to be displayed when [T] is pressed.



Experiment: IF01

TRT	Age	SPECIES	SITE	COUN.	SUR	HT	DIA
T01	12	ACACMANG	Hyd-bad	India	88	5.8	5.4
T04	12	ACACMANG	Hyd-bad	India	91	6.2	5.5
T07	12	ACACAURI	Hyd-bad	India	85	4.9	4.4
T10	12	ACACAURI	Hyd-bad	India	70	4.5	3.7
T13	12	LEUC_SPP	Hyd-bad	India	76	3.3	2.5
T16	12	LEUC_SPP	Hyd-bad	India	71	2.2	1.8
T01	19	ACACMANG	Hyd-bad	India	85	8.9	8.5
T04	19	ACACMANG	Hyd-bad	India	86	9.3	8.8
T07	19	ACACAURI	Hyd-bad	India	82	7.4	6.1
T10	19	ACACAURI	Hyd-bad	India	65	7.4	6.1

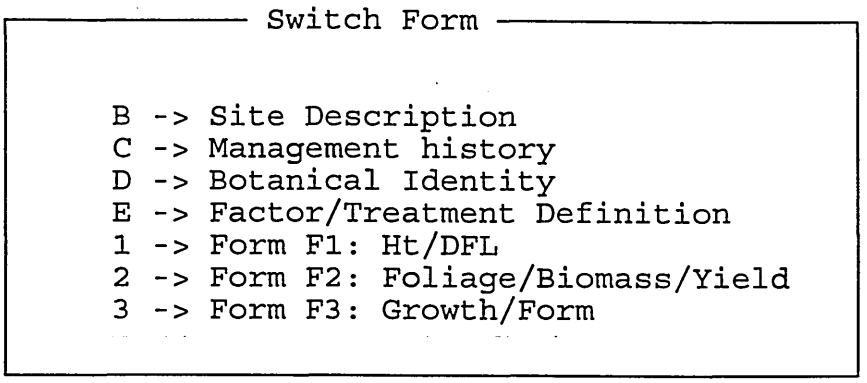
[Enter]=Display treatment [S]ave records to file
[F]orms menu
[ ] [ ] [ ] [ ] [PgUp] [PgDn]

[Esc]=Exit [F1]=Help [F2]=Print

Screen 22. Summary Display, Page 2.

Forms menu

You can also press [F] to display a menu of the other forms. By selecting a form from the menu the program will display information in that form that is linked to the current record highlighted in the summary display.



Screen 23. Forms Selection Window

NOTE: Information must be connected together with the ID's (i.e. Project ID, Site ID) when adding data. This allows the program to display information in other forms that are linked to the current record in form F.

Save records to file

Press [S] to save the performance records to a disk file for later use in searching or creating an analysis dataset. You will be prompted to enter a file name. Use descriptive file names using letters, digits, or the underline character. Begin the filename with a letter.

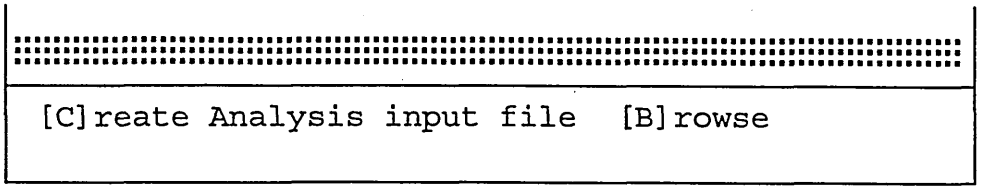
Exit

Press [Esc] return to the search screen.

DATA SETS FOR ANALYSIS

The Data Sets for Analysis option allows you to create analysis datasets containing selected variables from the master growth performance file, SUMF.DBF, or files you may have saved from the General Search option.

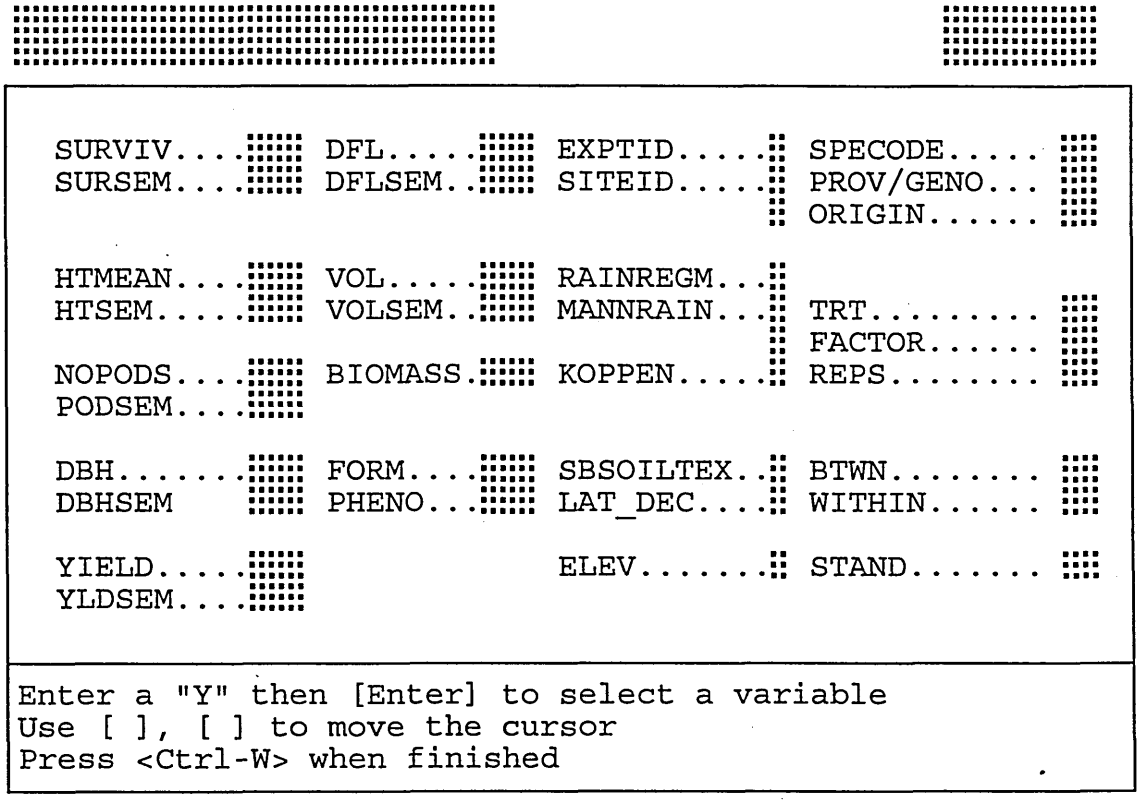
The operational keys available are [C] to create the data set by selecting the variables to be included, [B] to browse the current file, [F3] to change the current file, and [Esc] to return to the main menu. The current file status line displays the filename, the number of records, the number of experiments, and the date of the file.



Screen 24. Data Sets for Analysis

Create

Pressing [C] displays the input fields for variable selection. The initial values are 'N' for do not include. A variable is selected for data set by entering a 'Y' in its input field.



Screen 25. Variable Selection for Data Sets

The variables are grouped into response and explanatory variables. The first two columns are the response variables and the last two columns are the explanatory variables. Each variable, except FACTORS, if selected becomes a column in the dataset.

NOTE: At least one response variable must be selected.

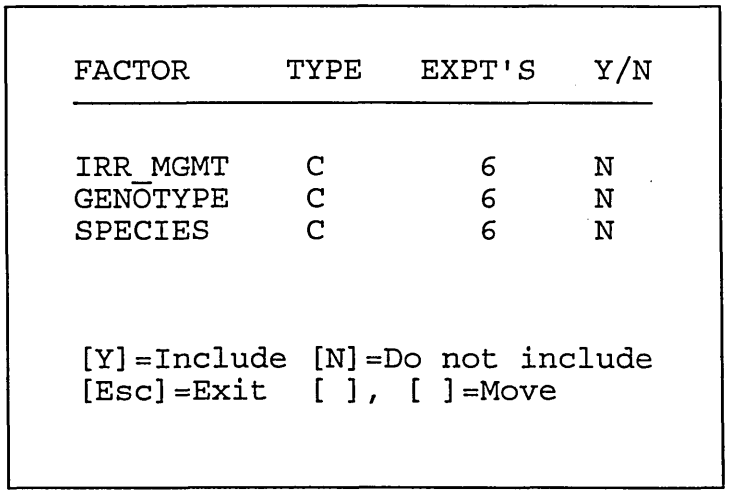
The description and type (numeric or categorical) of a variable can be seen by pressing [F1] when the cursor is at that variable's input field.

The explanatory variable AGE, if included, will be defined as a numeric variable. The explanatory variable AGE_C, if included, will contain the same data as AGE, but will be defined as a categorical or nominal variable.

The explanatory variable FACTORS, when selected, displays the Factors Selection Window.

This window shows a list of factors currently in the database, its type, the number of experiment that use the factor, and its selection flag. All the factors are initially not selected (they will not be included in the dataset). Enter a 'Y' beside the factor(s) you want to include in the dataset. Each factor becomes a column in the dataset.

NOTE: For performance records which do not use a particular factor, that column will be filled with a period, '.'. The period is used to indicate a missing value.



Screen 26. Factors Selection Window

After the variables are selected press <Ctrl-W> to begin creating the dataset. Messages and counters will be displayed as the dataset is being created. A prompt will appear for entering the dataset name.

NOTE: Use up to eight characters using letters, digits or the underline character. Use a letter as the first character and do not put spaces between the characters.

Browse
Pressing [B] displays the contents of the current file as they are stored.

LINKING INFORMATION TOGETHER

Information entered into the six data forms are linked together by using ID codes that you enter. These ID's will allow the program to display information from one form that is linked to another form. For example, if you are looking at site information in Form B (Site Description) after a search, the links (ID's) will allow the program to display information in Form A (Project Description) which is information about the project responsible for experiment(s) at the site described in Form B (as shown in figure 1).

The logical data structure illustrates the primary links (Project ID, Site ID, Experiment ID, and Seed origin) between the data files. Secondary links are used to support the search algorithm. The links (ID's, Codes, and Names) are stored in key fields in the database file records. Table 1 shows the key fields and the files (and corresponding data forms) which contain them.

SELECTION WINDOWS

Some of the information fields have a set of possible entries, for these fields a list can be displayed on the screen. The list is activated by pressing <Ctrl-Home>. This key combination is active when the message '<Ctrl-Home>=Pick List' appears at the bottom of the screen and 'Ctrl-Home' is displayed in bright type. To select an entry from the list move the highlight bar (using the up and down cursor keys) to the desired entry and press [Enter]. The select entry will be placed in the input field. [Esc] is used to exit. Data can be entered in these fields directly.

USING SAVED FILES

The save option in the general search option is used to store performance records found during a search for later use. The saved file (user file) can be used both in the general search and the data sets and analysis options. These options initially use as the current file SUMF.DBF, the master performance data. The [F3] key is used to change the current file to a previously saved file in both options.

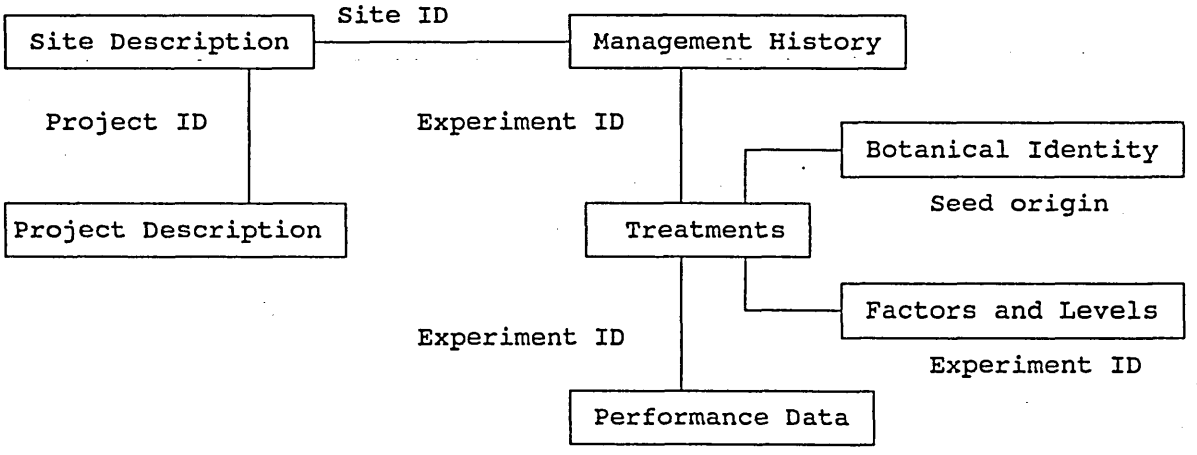


Figure 1. Logical Data Structure

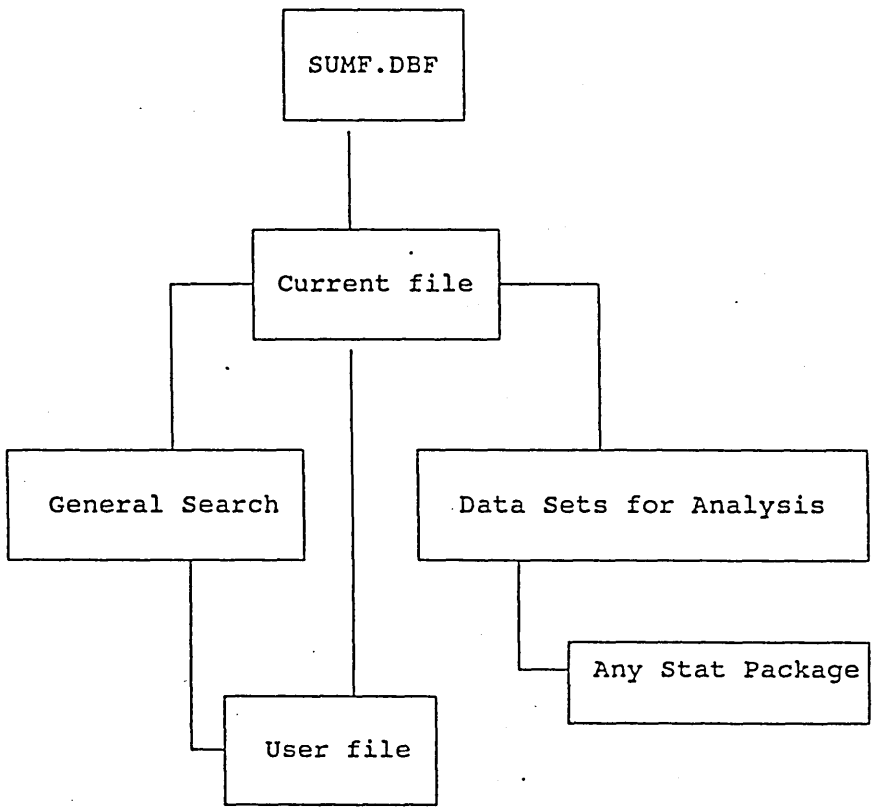


Figure 2. User file flow diagram.

Table 1. Key Fields

KEY FIELDS	DATABASE FILES									
	SumA	SumB	SumC	SumD	SumFac	Factors	SumLev	Levels	SumTrt	SumF
Project ID	Y	Y	-	-	-	-	-	-	-	-
Site ID	-	Y	Y	-	-	-	-	-	-	Y
Experiment ID	-	-	Y	-	Y	-	Y	-	Y	Y
Species/Crop Code	-	-	-	Y	-	-	-	-	-	Y
Seed Origin	-	-	-	Y	-	-	-	-	-	Y
Factor Name	-	-	-	-	Y	Y	Y	Y	-	-
Level Name	-	-	-	-	-	-	Y	Y	Y	-
Treatment Name	-	-	-	-	-	-	-	-	Y	Y
	A	B	C	D	E-1	E-1	E-1	E-1	E-2	F
	DATA FORMS									

Table 2. Search Fields

Field	1	2	3	4
Species/Crop	-	-	-	-
Genotype	Y	-	Y	Y
Seed Origin	Y	-	Y	Y
Rainfall Regime	-	Y	Y	Y
Mean Annual Rainfall	-	Y	Y	Y
Mean Annual Temperature	-	Y	Y	Y
Climatic Code	-	Y	Y	Y
Subsoil texture	-	Y	Y	Y
Age/Date of planting	-	-	Y	Y
Factor 1	-	-	-	Y
Factor 2	-	-	-	Y
Factor 3	-	-	-	Y
Levels of Factor 1	-	-	-	Y

DATA DICTIONARY

FILE NAME:SumA.DBF — DESC.:Project Description

FIELD	TYPE	LEN	DESC/REMARK
PROJID	C	8	Project name/ID, must be unique per record, stored ltrim(), NOT case sensitive, NO embedded blanks, alphanumeric and underscore, starting with nonnumeric.
NAME	C	15	Proponent or Manager name
ADDR1	C	40	Address line 1
ADDR2	C	40	Address line 2
ADDR3	C	40	Address line 3
ADDR4	C	40	Address line 4
ADDR5	C	40	Address line 5

INDEX NAME:	INDEX STRING:
AProjID.NDX	upper(PROJID)

FILE NAME:SumB.DBF — DESC.: Site Description

FIELD	TYPE	LEN	DESC/REM
SITEID	C	8	Local name/number of trial site, must be unique per record, store ltrim(), NOT case sensitive, NO embedded blanks, alphanumeric and underscore, starting with nonnumeric.
SITENAME	C	15	Local site name
STATE	C	2	State abbreviation
COUNTRY	C	4	Country abbreviation
PROJID	C	8	Project name/id (from form A), store ltrim(), NOT case sensitive, NO embedded blanks, alphanumeric and underscore, starting with nonnumeric
LAT_DEG	C	2	Degrees latitude
LAT_MIN	C	2	Minutes latitude
LAT_NS	C	1	North/South latitude
LON_DEG	C	3	Degrees longitude
LON_MIN	C	2	Minutes longitude
LON_EW	C	1	East/West longitude
ELEV	C	4	Elevation(m)
RAINREGM	N	2	Rainfall regime
MANNRAIN	C	5	Mean annual rainfall (mm)
LENDRY	C	2	Length of dry season (# months <40mm)
MANNTEMP	C	4	Mean Annual temp (c)
MMINTEMP	C	4	Mean daily min temp coldest month (c)
MMAXTEMP	C	4	Mean daily max temp hottest month (c)
ABSMINTEMP	C	4	Absolute min temperature (c)
SRSOILTEX	N	1	Surface soil texture
SBSOILTEX	N	1	Subsurface soil texture
SRSOILCLR	N	1	Surface soil color
AVEDEPTAB	N	1	Average depth to water table
SOILMOISP	N	1	Soil moisture at planting
SOILMOISA	N	1	Soil moisture after planting
SRSOILPH	N	1	Surface soil pH description
SBSOILPH	N	1	Subsurface soil pH description
SRSOILSAL	N	1	Surface soil salinity
SBSOILSAL	N	1	Subsurface soil salinity

INDE NAME:	INDEX STRING:
BSteId.NDX	upper(SITEID)
BRnReg.NDX	RAINREGM
BAnnRn.NDX	MANNRAIN
BTemp.NDX	MANNTEMP
BSTex.NDX	SBSOILTEX

FILE NAME:SumC.DBF DESC:Management History

FIELD	TYPE	LEN	DESC/REM
EXPTID	C	8	Experiment ID, must be unique per record, stored ltrim(), NOT case sensitive, NO embedded blanks, alphanumeric and underscore, starting with nonnumeric.
SITEID	C	8	Site ID (from form B), stored ltrim(), NOT case sensitive, NO embedded blanks, alphanumeric and underscore, starting with nonnumeric.
YM_PLANT	C	4	Year and month Planted
REP	C	2	Number of replicates
NUMTRT	C	2	Number of treatments
NATTRT1	C	2	1st Nature of treatment
NATTRT2	C	1	2nd Nature of treatment
NATTRT3	C	1	3rd Nature of treatment
PLOTBLK	C	2	Number of plots per block
DIMX	C	4	X dimension of plot (m)
DIMY	C	4	Y dimension of plot
BTWN	C	4	Spacing between rows
WITHIN	C	4	Spacing within rows
FERT	C	1	Fertilier (Y/N)
FERTC	C	30	Fertilizer comment
IRRI	C	1	Irrigation (Y/N)
IRRIC	C	30	Irrigation comment
INOC	C	1	Inoculum (Y/N)
INOCC	C	30	Inoculum comment
WEED	C	1	Weed control (Y/N)
WEEDC	C	30	Weed control comment

INDEX NAME	INDEX STRING
CExptID.NDX	upper(EXPTID)
CBtwn.NDX	BTWN
CWthin.NDX	WITHIN

FILE NAME:SumD.DBF DESC:Botanical Identity

FIELD	TYPE	LEN	DESC/REM
SEEDORIGIN	C	8	Seed origin details of supplier's name, address etc.
SPECODE	C	8	Species code, 4,4, stored ltrim(), NOT case sensitive, NO embedded blanks, alphanumeric, starting with nonnumeric.
GENOTYPE	C	8	Genotype names
COUNTRY	C	4	Country name
STATE	C	4	State or Territory
LAT_DEG	C	2	Degrees latitude
LAT_MIN	C	2	Minutes latitude
LAT_NS	C	2	North/South latitude
LON_DEG	C	2	Degrees longitude
LON_MIN	C	2	Minutes longitude
LONG_EW	C	1	East/West longitude
ELEV	C	4	Elevation
SOURCE	C	30	Species information.

INDEX NAME	INDEX STRING
DSLot.NDX	upper (SEEDLOT)
DProv.NDX	upper (GENOTYPE)

FILE NAME:SumFac.DBF — DESC. Experiment factors

FIELD	TYPE	LEN	DESC/REM
EXPTID	C	8	Experiment ID (from form C), stored ltrim(), NOT case sensitive, NO embedded blanks, alphanumeric and underscore, starting nonnumeric.
FAC1	C	8	Factor 1 (from factors.dbf), stored ltrim(). alphanumeric and underscore, starting with nonnumeric
FAC2	C	8	Factor 2 name (from factors.dbf), stored ltrim(), alphanumeric and underscore, starting with nonumeric.
FAC3	C	8	Factor 3 name (from factors.dbf), stored ltrim(), alphanumeric and underscore starting with nonnumeric.

NOTE: No duplicate factor per experiment

INDEX NAME	INDEX STRING
E1AExpt.NDX	upper(EXPTID)
E1AFac1.NDX	upper(FAC1)
E1AFac2.NDX	upper(FAC2)
E1AFac3.NDX	upper(FAC3)

FILE NAME:Factors.DBF — DESC: Factor name and description

FIELD	TYPE	LEN	DESC/REM
FACNAME	C	8	Factor name must be unique per record, stored ltrim(), NOT case sensitive, NO embedded blanks, alphanumeric and underscore, starting with non-numeric.
FACDESC	C	40	Factor description

INDEX NAME:	INDEX STRING
Factors.NDX	upper(FACNAME)

FILE NAME:SumLev.DBF — DESC.:Experiment factor levels

FIELD	TYPE	LEN	DESC/REM
EXPTID	C	8	Experiment ID (from form C), stored ltrim(), alphanumeric and underscore, starting with nonnumeric.
FACNAME	C	8	Factor name (from SumFac.DBF), alphanumeric and underscore, starting with nonnumeric.
LEVNAME	C	8	Level name (form levels.dbf), alphnumeric and underscore, starting with nonnumeric.

NOTE: upper(EXPTID) + upper(FACNAME) is unique per record

INDEX NAME:	INDEX STRING
E1BExpt.NDX	upper(EXPTID) + upper(FACNAME)
E1BFac.NDX	upper(FACNAME)
E1BLev.NDX	upper(LEVNAME)

FILE NAME:SumF.DBF DESC.: Performance data

FIELD	TYPE	LEN	DESC/REM
EXPTID	C	8	Experiment ID (from form C), alphanumeric and underscore, starting with nonnumeric.
TRT	C	8	Treatment name/ID (from SumTrt), alphanumeric, starting with nonnumeric.
SITEID	C	8	Local name/number of trial Site name.
SPECODE	C	8	Species code 4,4 store ltrim(), alphanumeric, starting with nonnumeric.
AGE/DATE	C	8	Age of plant or date of planting
REPS	C	2	Number of replications.
SURVIV	C	3	Survival (%)
SURSEM	C	3	Survival standard error of mean
HTMEAN	C	4	Height mean
HTSEM	C	3	Height stardard error of mean
DFLMEAN	C	4	Days to flowering mean
DFLSEM	C	3	DFL standard error of mean
DIAMEAN	C	4	Diameter mean
DIASEM	C	3	Diameter Standard error of mean
FOLIAGE	C	4	Foliage mean
FOLSEM	C	3	Foliage standard error of mean
BIOMASS	C	4	Biomass mean
BIOSEM	C	4	Biomass standard error of mean
YIELD	C	4	Yield mean
YLDSEM	C	3	Yield standard error of mean
STAND	C	4	Stand density at measure
GROWTH	N	1	Growth relative to other taxa
FORM	N	1	Plant form

NOTE: Exptid + Age + upper(Trt) is unique per record

INDEX NAME:]	INDEX STRING
FExTrt.NDX	upper(EXPTID)+AGE+upper(TRT)
FSiteId.NDX	upper(SITEID)
FAge.NDX	Age + upper(EXPTID)
FSpec.NDX	upper(SPECODE)+upper(EXPTID)+Age+TRT
FSLot.NDX	upper(SEEDLOT)+upper(EXPTID)+Age+TRT

DATA DICTIONARY

FILE NAME:Levels.DBF — DESC: Level name and description

FIELD	TYPE	LEN	DESC/REM
FACNAME	C	8	Factor name (from Factors.dbf), stored ltrim(), alphanumeric and underscore, starting with nonnumeric.
LEVNAME	C	8	Level name stored ltrim(), alphanumeric and underscore, starting with nonnumeric.
LEVDESC	C	40	Level description.

NOTE: upper(FACNAME) + upper(LEVNAME) is unique per record.

INDEX NAME:	INDEX STRING
LevelsF.NDX	upper(FACNAME) + upper (LEVNAME)
Levels.NDX	upper(LEVNAME)

FILE NAME:SumTrt.DBF — DESC. Treatment definition

FIELD	TYPE	LEN	DESC/REM
EXPTID	C	8	Experiment ID (from form C), stored ltrim(), alphanumeric and underscore, starting with nonnumeric.
TRT	C	8	Treatment code/name, stored ltrim(), alphanumeric and underscore, starting with nonnumeric.
LEV1	C	8	Level 1 (from sumlev.dbf), stored ltrim(), alphanumeric and underscore, starting with nonnumeric.
LEV2	C	8	Level 2 (from sumlev.dbf), stored ltrim(), alphanumeric and underscore, starting with nonnumeric.
LEV3	C	8	Level 3 (from sumlev.dbf), stored ltrim(), alphanumeric and underscore, starting with nonnumeric.
SEEDORIGIN	C	8	Seed origion, and suppliers code or name.

NOTE: upper(EXPTID) + upper(TRT) is unique per record

INDEX NAME:	INDEX STRING
E2ExpTrt.NDX	upper(EXPTID)+upper(TRT)